

# NAVIGATING THE AI REVOLUTION: A BUSINESS LEADER'S GUIDE

---

The Defining Developments in AI  
Software, Safety and Scale

FIVE STARTUPS TO  
WATCH IN 2025

# Contents

<b>Executive Summary</b>	<u>04</u>
About the Authors	<u>08</u>
<b>Methodology</b>	<u>09</u>
Parker & Lawrence's 2024 Research in Numbers	<u>10</u>
Scope	<u>11</u>
<b>Key Trends</b>	<u>14</u>
Software: Adoption and Delivery	<u>14</u>
Optimised Architectures	<u>17</u>
Retrieval-Augmented Generation	<u>17</u>
Ensemble Models	<u>22</u>
Custom Models	<u>25</u>
Beyond Chatbots	<u>27</u>
Agents	<u>28</u>
Semantic Layers	<u>36</u>
Safety: Risk and Compliance	<u>39</u>
Defining Risk	<u>40</u>
Key Risks and Regulations	<u>40</u>
Measuring Risk	<u>48</u>
Evaluations and Testing	<u>48</u>
Addressing Risk	<u>52</u>
Governance and Tooling	<u>52</u>
Scale: Infrastructure and Energy	<u>56</u>
Infrastructure	<u>57</u>
Data Centers	<u>57</u>
Energy	<u>65</u>
Powering AI	<u>65</u>
<b>Conclusion</b>	<u>71</u>
Services as a Bridge	<u>72</u>
Stay Ahead in The AI Revolution	<u>73</u>
AI Revolution Datasheet	<u>74</u>



# Executive Summary

## Welcome to the early stages of our AI revolution

We are in the very early stages of the AI revolution. The scale and scope of its impact—on industries, economies, and societies—remain uncertain. What kind of revolution this will be, and what long-term implications it will have, is still up for us to decide. What is certain, however, is that AI represents a true revolution. While a healthy degree of skepticism remains, early AI adoption, impacts, plans and projections tell a compelling story (jump to the [AI Revolution Datasheet<sup>1</sup>](#).)

The positions of investors, institutions, and governments regarding AI are becoming increasingly clear. Each group is aligning their strategies around the transformative potential of the technology. Yet, the precise course AI will take depends on the timing and nature of technical breakthroughs—and the organisations or individuals driving and regulating them.

Consider an extreme and highly improbable scenario: the achievement of true Artificial General Intelligence (AGI) in 2025, perhaps by OpenAI. Such a development would disrupt the positions of all stakeholders overnight. Governments would scramble to respond, likely introducing emergency policies to address societal and security implications. Investors would reallocate funds, seeking to capitalise on or mitigate risks associated with such a paradigm shift. Institutions would face immediate operational and strategic adjustments, while individuals like Sam Altman, at the helm of these advancements, would emerge as a central figure in reshaping global norms.

This hypothetical scenario underscores a broader truth: the illusion of control and stability dissolves upon each technological step change. As history has shown, revolutions in technology are rarely linear or predictable.

### The Curse of the Commentator

In the middle of our research, a technical breakthrough did indeed occur: DeepSeek, a Chinese AI startup, released its R1 model on January 20, 2025. This open-source AI model demonstrated reasoning capabilities comparable to leading western models but was developed at a fraction of the cost. DeepSeek's R1 quickly became the most-downloaded free app on the U.S. iOS App Store, surpassing ChatGPT. This achievement led to a substantial market reaction, with Nvidia's stock plummeting by up to 18% and over \$1 trillion being wiped off American tech stocks. Eerily, over the same period, Sam Altman was photographed standing alongside U.S. President Donald Trump during the announcement of the Stargate Project—a \$500 billion investment in AI infrastructure.



# HOW TO THRIVE IN THE AI REVOLUTION



## Understand, don't overfit

Track developments in AI with a clear eye on their relevance to your domain, weighing their true potential against their immediate appeal. Many companies overreacted to the step change in LLM capabilities, rapidly pivoting on features, products, and even entire strategies. The result? A flood of text-generation tools that have already started losing market share.<sup>2</sup> Avoid chasing fleeting trends; instead, focus on strategic adoption that aligns with lasting value. First-mover advantage is often overstated—and last-mover advantage forgotten altogether.

## Start with why, not AI

Rapid deployment in a fast-evolving market may create short-term wins, but long-term success depends on understanding deeper trends and responding strategically. Successful projects anchor innovation on the fundamental problems and needs of their domain, which are often more stable than the technologies addressing them. Building impressive technology is exciting, but building technology that solves enduring human problems is transformative. Understanding your user is the ultimate competitive advantage, enabling you to effectively leverage emerging technologies.

## Build Solid flexible foundations

The ground is shifting beneath us, and rigid strategies or systems will inevitably crumble. Flexibility is critical—whether in governance frameworks, development methodologies, or investment strategies. A dynamic and well-structured regulatory regime can adapt quickly to breakthroughs, agile teams are better equipped to pivot as needed, and diversified investors are more likely to capture value from emerging opportunities. Building with flexibility in mind ensures that you can evolve alongside technology, rather than being disrupted by it.



## In other words, become antifragile: That which gains from disorder.

The purpose of our research is to reduce uncertainty—at least for today, if not for tomorrow. This report highlights key statistics, trends, and standout companies from our research, which are set to become increasingly relevant in 2025.

The evolution of AI in 2025 will be defined by significant shifts across **software, safety, and scale**. To help business leaders navigate this transformation, we take a practical, non-technical approach, connecting technical developments to AI's real-world impacts. Each trend is broken down into three levels of granularity—**What You Need to Know**, **What You Should Know**, and **What Experts Are Talking About**—guiding readers from foundational understanding to deeper industry insights. The result is an actionable, structured framework for decision-makers looking to stay ahead.

## ► Software

AI adoption is accelerating, but impact remains uneven—while leaders see up to 50% higher revenue growth, many struggle to justify ROI. The next phase of AI software is defined by modular, enterprise-grounded systems that enhance accessibility, automation, and integration.

A few key trends are reshaping adoption: Retrieval-Augmented Generation (RAG) is mitigating hallucinations by grounding AI in real-time data, ensemble models improve accuracy and reliability, and custom AI models offer domain-specific performance while reducing dependency on third-party providers. AI agents are moving beyond chat to become proactive digital co-workers, while semantic layers structure data for more meaningful insights.

## ► Safety

AI's rise brings growing risks—bias, misinformation, security threats, and regulatory uncertainty. Fortune 500 AI risk disclosures have surged 473.5% since 2022, yet many organisations lack structured governance.

### ● Defining risk:

AI regulation is evolving but fragmented, with the EU enforcing strict compliance while the U.S. and U.K. take a flexible approach. Businesses must navigate these complexities.

### ● Measuring risk:

AI evaluations remain inconsistent, with many skipping critical testing. Automated validation and fairness audits are becoming essential safeguards.

### ● Addressing risk:

Strong governance and compliance tools are key. Industry leaders are implementing internal safeguards, recognising that AI safety can't rely on regulation alone.



## ► Scale

AI adoption is accelerating, but scaling it is becoming increasingly complex. **83% of organisations expect their AI workloads to grow in the next two years**, yet scaling isn't just about investment—it's constrained by energy, infrastructure, and computing power.

- **Infrastructure:**

Data center demand is surging, with hyperscalers like Microsoft and Google investing billions to expand capacity. Governments are stepping in too, with initiatives like the \$500 billion U.S. Stargate Project. Still, supply struggles to meet demand, with vacancy rates in key hubs below 1%.

- **Energy:**

AI's energy consumption is skyrocketing, expected to drive 8% of U.S. electricity demand by 2030. Power constraints are pushing investment in renewables and nuclear energy, yet fossil fuels remain central to AI's expansion.

- **Efficiency:**

The future of AI scale will be defined by technological optimisations. Specialised AI chips, smarter workload management, and improved cooling systems are becoming essential for reducing costs and energy consumption at scale.



## About the Authors

Parker & Lawrence is a boutique market research firm specialising in AI, with a particular focus on financial services, risk, and compliance. We are commissioned by investors, enterprises, technology vendors, and governments to provide market insights through primary research—including expert interviews and large-scale surveys—through white papers, technical product marketing and other thought leadership materials. This work supports **our mission: to enable the responsible adoption of emerging technologies.**

**This paper was authored by its two founders:**

### Nathan Parker

Nathan is a thought leader in RegTech, FinTech, and Web3, with a track record of delivering high-impact research for global technology vendors and regulators. His expertise has been instrumental in helping RiskTech and RegTech firms develop and launch innovative solutions, ensuring market success both domestically and internationally.

### Michael Lawrence

Michael is a technology researcher specialising in AI, risk, and compliance. With hands-on experience in building technology, he has worked on machine learning and large language models in both financial services and government. Since 2017, he has focused on AI market research, advising major regulators and financial institutions on technology strategies, serving as a Product Manager for a digital marketplace for B2B software solutions, and producing extensive thought leadership.



# Methodology

The insights in this report were derived from **four main sources**:

- Passively developed through our extensive, ongoing market research. Check out our [2024 research in numbers](#).
- A targeted analysis of our startup database, focusing on companies funded in 2024. Check out the [scope](#) of our analysis.
- A synthesis of 31 research reports, representing perspectives from a total of **34,173 survey responses** (all referenced within this paper).
- In-depth research interviews with selected high-potential AI startups.

While much of our analysis was focused on AI companies funded in 2024, our selection of startups was biased towards those funded in 2023—who have had over a year to refine their product and establish positioning—who are set to gain from the trends identified in this paper and who we believe are poised for substantial growth in 2025.

## 5 startups to look out for in 2025



[Numbers Station](#) empowers users to automate analytics workflows via an ecosystem of specialist AI Agents, all in natural language.



[Akooda](#) is an AI-powered enterprise search solution, connecting people, projects, and insights in real time across all company tools.



[Eve Legal](#)'s AI-driven platform streamlines case workflows from intake to resolution, enabling firms to handle more cases with existing staff.



[Harmonya](#) aligns, enriches, analyses and operationalises consumer packaged goods data (CPG) at scale, driving growth through insight.



[SplxAI](#) automates AI red teaming to detect and mitigate generative AI threats in real time.



## The Curse of the Commentator... Again.

Already, these selected startups are proving our predictions correct. After performing research interviews in late 2024, they are each already making significant strides in 2025. In Q1 alone, three have closed major funding rounds:

- Eve Legal raised an unprecedented **\$47 million Series A**, led by Andreessen Horowitz.
- Harmonya secured their **Series B**, led by dunnhumby Ventures, specialists in AI, data, and analytics startups in the retail and CPG ecosystem.
- SplxAI announced it has closed **\$7M** in seed funding led by LAUNCHHub Ventures with participation from Rain Capital, Inovo, Runtime Ventures, DNV Ventures and South Central Ventures.

## Parker & Lawrence's 2024 Research in Numbers

On top of client projects, research is a core part of our day-to-day business. We focus on early-stage companies, particularly those at Series A and earlier—where innovation is most dynamic, risks are highest, and truly "emerging" technologies take shape. In 2024, our research was extensive in this space:



Built our AI Startups database to  
**2879** companies



**61** Product demos



Quantitative analysis of  
**1484** AI startups



**2** unique Generative AI (GenAI) Taxonomies: capabilities and risks.

- Used to map **14** risk and compliance use cases.
- With help from **17 expert contributors** in our previous report.<sup>3</sup>



Qualitative analysis of  
**223** AI startups

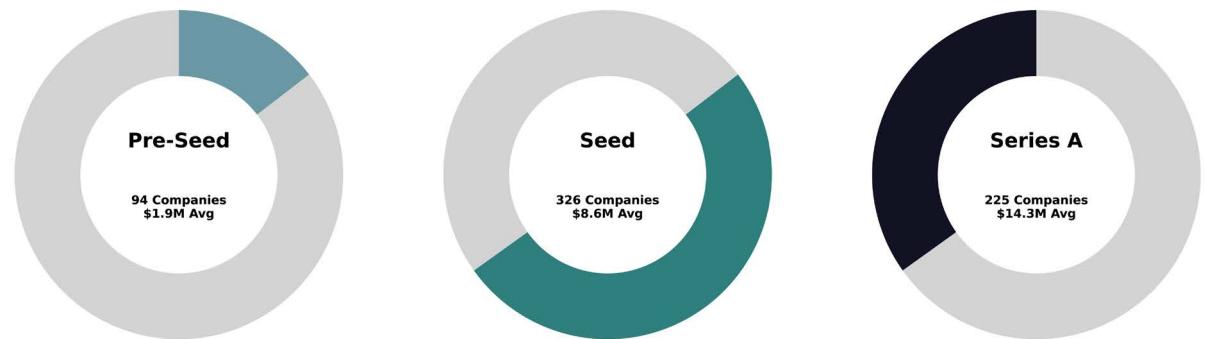
# Scope

For the purposes of this report, we also took a closer at the AI startups in our database that were funded in 2024:

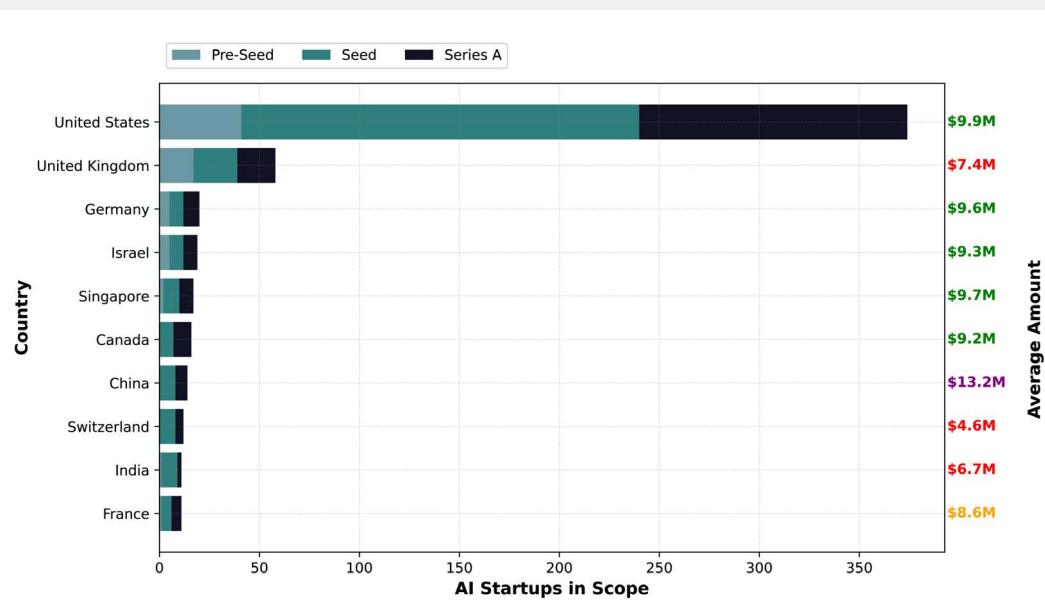
**645**  
Companies

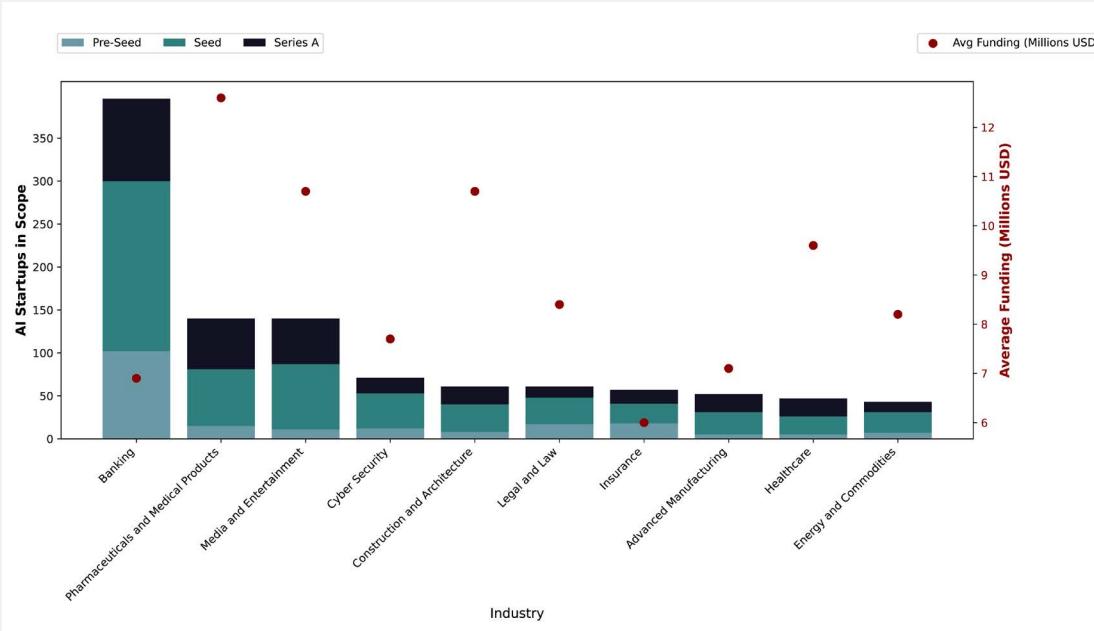
**\$5.85B**  
Raised

Across 3 Funding Stages:



In 41 countries, including:



**Covering 19 Industries, Including:**

## The AI Opportunity in Banking

The strong banking focus in our data aligns with broader research from McKinsey, which highlights the immense AI opportunity in financial services. As banks face slowing revenue growth, rising compliance costs, and increased competition from fintechs and neobanks, AI powered automation and predictive analytics offer a critical advantage in cost efficiency and margin protection.

AI is already transforming loan underwriting, risk modeling, and software development, with some banks reporting a 40% increase in coding productivity. In retail banking, AI is driving hyper-personalised customer engagement, fueling revenue through smarter financial planning and credit insights.

McKinsey points out that the banks excelling with AI have resisted the urge to quickly build narrow chatbots, and are instead integrating AI across core functions—including risk, compliance, sales, and customer service—to unlock its full potential.<sup>4</sup>

Citi also finds that the financial services industry is among the fastest adopters of AI, behind only telco & media.<sup>5</sup> And they point out the potential for cutting-edge AI developments, particularly AI agents, to transform the industry further.

GenAI alone is expected to generate between \$200 billion and \$340 billion in new value for banks,<sup>6</sup> and a growing number of tools are emerging to help the industry realise that value. A few companies have caught our attention:

- **Theia Insights** delivers AI-driven investment analytics, using a self-learning knowledge graph to enhance industry classification, thematic investing, and portfolio risk assessment.<sup>7</sup>
- **Sibli** leverages AI to extract insights from unstructured data, equipping institutional investors with decision-making tools that streamline research, automate reporting, and uncover high-conviction signals.<sup>8</sup>
- **Brightwave** accelerates financial research with an AI-powered diligence platform that reads every line of filings, transcripts, and contracts to surface critical insights. By synthesizing vast amounts of financial data, Brightwave helps investors and analysts make high-conviction decisions with speed and confidence.<sup>9</sup>



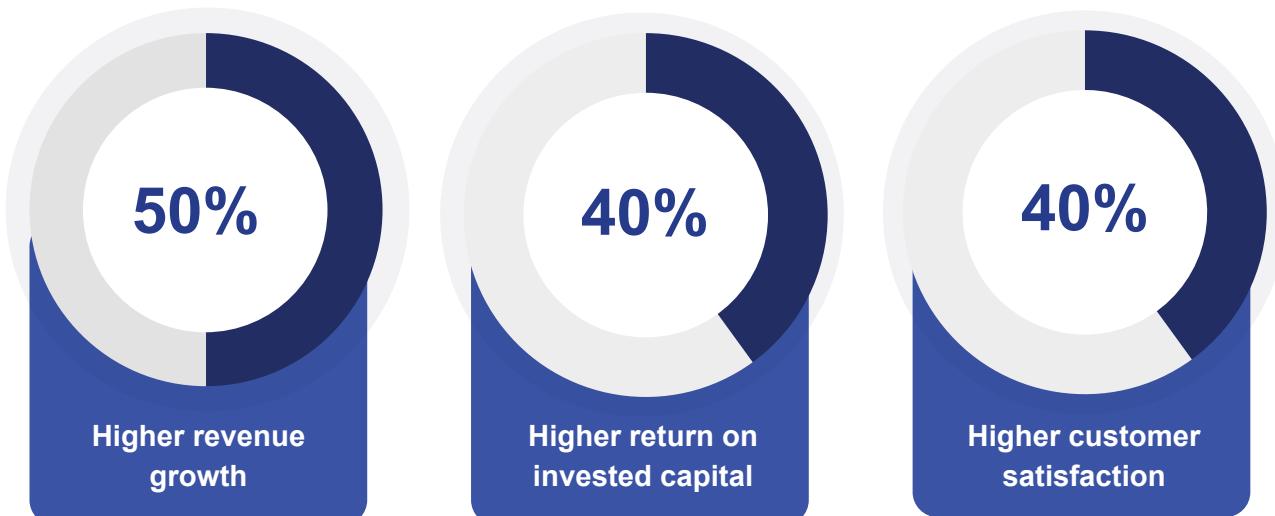
# Key Trends

## Software: Adoption and Delivery

### Background

Revolutions were never easy. While we have recently witnessed a step change in AI's capabilities—and a surge in adoption and impact—these early stages are defined as much by struggle as they are by excitement.

BCG's October 2024 research found that 74% of companies are not yet achieving value at scale with AI. Organisations leading in AI adoption, however, are realising enormous benefits:



A few key behaviours make these AI leaders different :<sup>10</sup>

- **Fast GenAI Adoption:**

Leaders rapidly integrate GenAI for content creation, reasoning, and automation, leveraging advanced capabilities.

- **Bold AI Ambitions:**

They use AI beyond cost-cutting, driving core business innovation they allocate more than 80% of their AI investments in reshaping core business functions and creating new products.

- **Strategic Focus:**

Instead of spreading efforts thin, they prioritise fewer, high-impact AI initiatives: 3.5 prioritised use cases on average, compared with 6.1 in lagging firms.



At the same time, S&P Global reported on a recent survey in which 49% of respondents claim GenAI use cases are not delivering the value anticipated. However, they also note that organisations are maturing in their selection of AI systems:



"An important shift is underway whereby organisations are moving away from standalone LLMs toward more modular GenAI-based systems that are grounded in an enterprise's data and have governance policies at their core"

— S&P Global, 2024 <sup>11</sup>



While IBM's Global AI Adoption Index includes a survey of enterprise IT Professionals who list their key drivers of adoption:<sup>12</sup>

Advances in AI tools that improve accessibility.

45%

### Delivery Matters

Look out for visualisations, guided workflows and low- and no-code tools.

Need to reduce costs and automate key processes.

42%

### Automation Matters

Look out for autonomous agents specialised and working together.

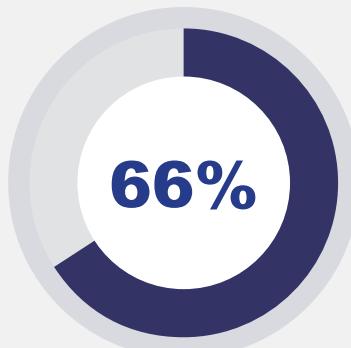
Increasing AI embedded in off-the-shelf applications.

37%

### Integration Matters

Look out for RAG tools, seamlessly accessing relevant, unique data.

We have been pleased to find that even early stage AI vendors are building with these factors in mind. Though the challenge isn't just building advanced solutions—it's ensuring businesses understand their value. Many organisations **struggle to justify AI investments**, with:

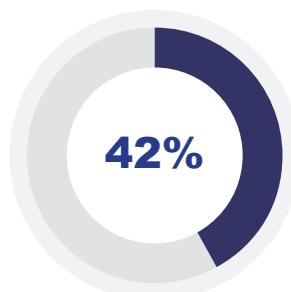


Citing difficulty establishing  
ROI on AI opportunities

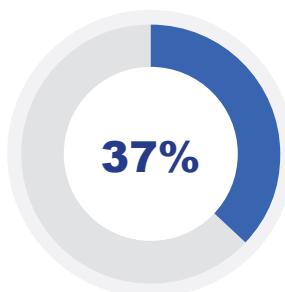


Struggling to make a business  
case for scaling AI initiatives

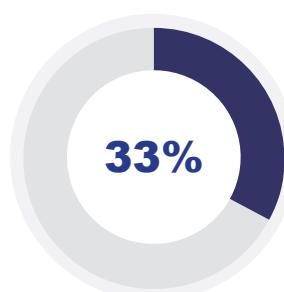
This underscores a broader issue: **many executives still lack a clear understanding of AI's business potential**. Key barriers include:



Lack a clear AI case  
for change



Struggle to reimagine  
workflows and processes



Cite insufficient AI  
literacy

There is more work to be done in educating the market with thought leadership that interests and speaks to business leaders. And vendors should be prepared to back it up with measurable impact when buyers come calling.

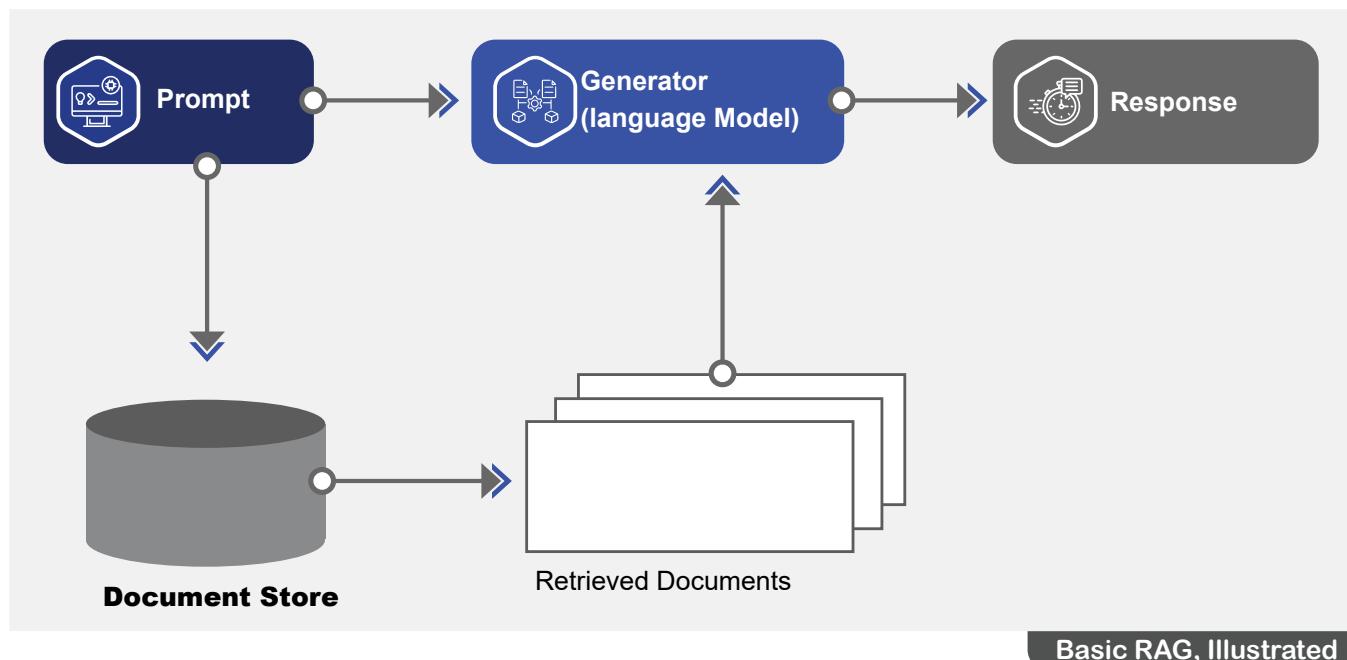
The remainder of this section details the key trends that are shaping AI adoption and delivery—from advancements in optimised architectures and autonomous agents to the evolving role of AI via APIs and visually-derived insights. These trends highlight how enterprises are refining their AI strategies, moving beyond experimentation toward scalable, high-impact implementations that drive real business value.

# Optimised Architectures

## Retrieval-Augmented Generation

### What You Need To Know: What is Retrieval-Augmented Generation?

RAG is an AI architecture designed to improve the accuracy and relevance of GenAI models by **incorporating real-world data from trusted sources**. Unlike standard AI models that generate responses based solely on their training data, RAG dynamically retrieves information from external sources (such as live web search) or internal knowledge bases (including company documents) before generating answers.



### What You Should Know: Why is RAG Popular?

RAG has become the go-to architecture for large enterprises due to its ability to overcome key limitations of standard GenAI models. By retrieving and grounding responses in trusted knowledge sources, **RAG mitigates risks** such as:





### Outdated Information

Allows for real-time access to the most relevant and current data instead of relying solely on static model training.



### Explainability & Compliance

Improves transparency by referencing sourced data, which is critical in regulated industries.

For large organisations handling vast amounts of proprietary data, RAG is particularly appealing. Enterprises often require deep domain and company specificity in AI-generated outputs. Additionally, many operate under stringent governance frameworks and AI risk measures, making RAG an attractive solution to keep AI-generated insights controlled, compliant, and relevant.

This view is supported by other research:

- ▶ McKinsey estimates that **25-40% of enterprises will adopt RAG by 2030**<sup>13</sup>, and that RAG adoption will be the largest driver of solid state drives (SSDs).<sup>14</sup>
- ▶ Vellum found that almost **60% of its 1250 respondents (AI Developers) are building RAG models**, noting that RAG remains the go-to solution—especially for enterprises.<sup>15</sup>
- ▶ Grand View Research found that the **RAG market was already worth over \$1 billion in 2023** and projects that it will grow at a CAGR of 44.7% from 2024 to 2030.<sup>16</sup>
- ▶ S&P's 451 Research found that **87% of enterprise leaders believe that RAG architectures enhance LLMs**.<sup>17</sup>

## Key Use Cases

### Real-Time Market & Competitive Intelligence:

Using RAG to retrieve live market trends, financial reports, and competitor updates for up-to-date decision-making.

### Customer Support & AI-Driven Chatbots:

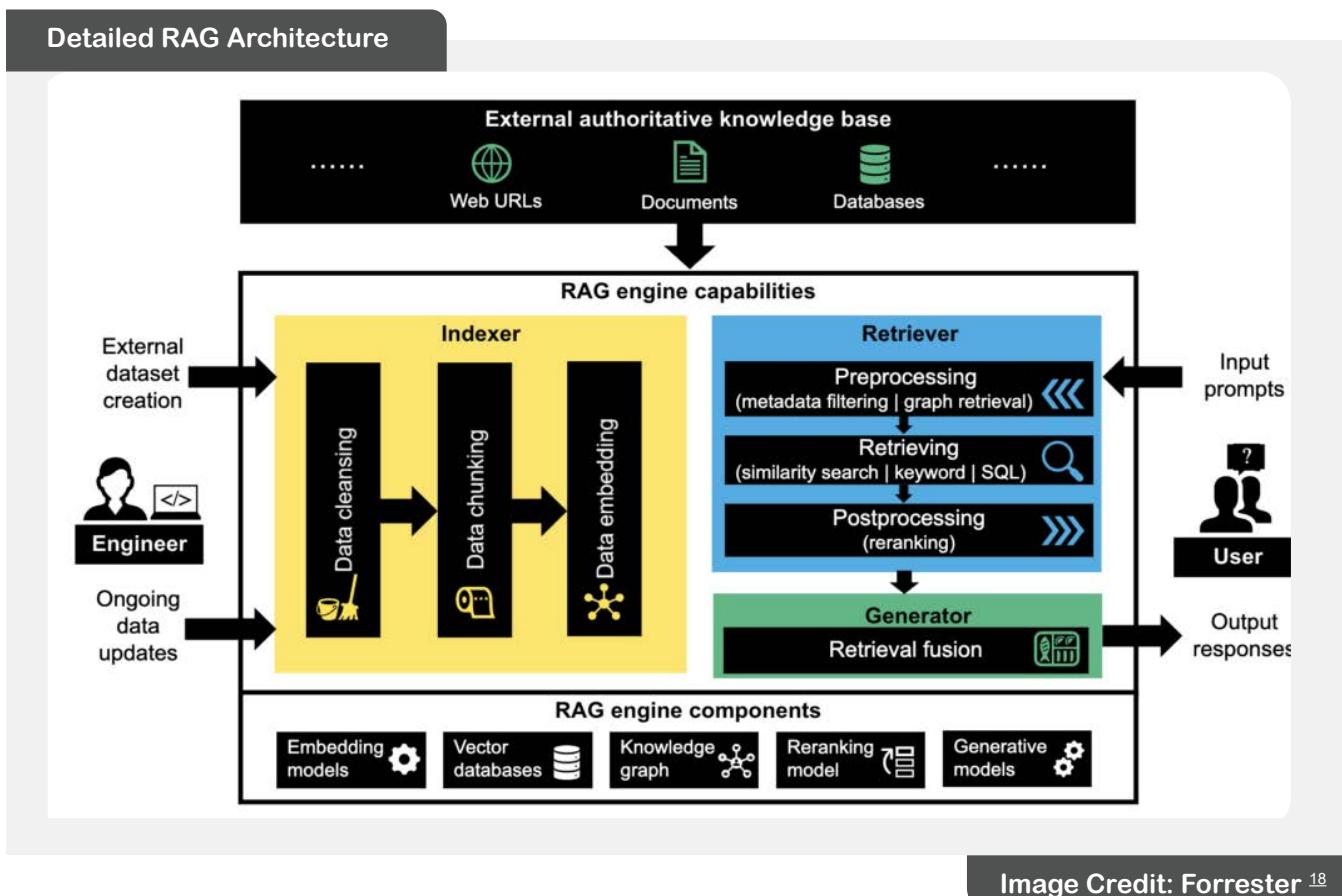
Allowing chatbots to reference internal FAQs, policies, and past interactions to provide accurate, company-specific responses.

### Enterprise Knowledge Management & Search:

Enabling employees to use AI to retrieve insights from internal documents, wikis, and databases, improving efficiency and decision-making.

<sup>14</sup> SSDs are high-speed storage devices used in GenAI infrastructure to store and quickly retrieve large datasets, including vector databases for RAG and AI model inferencing.

## What Experts Are Working On: Advanced RAG



While RAG mitigates hallucinations and improves contextual relevance, questions remain about its robustness:

- Does it truly prevent hallucinations?
- How reliable is the data being referenced?
- What measures ensure AI responses remain explainable and compliant?

Researchers and vendors are continuously refining its components to enhance accuracy, security, and efficiency:

#### Indexer – The Organiser

- Prepares and structures data so it's ready for retrieval.
- Cleans, chunks, and embeds information from documents, databases, or other sources.
- Ensures only relevant, high-quality data is available for AI to reference.

#### Retriever – The Researcher

- Finds the most relevant information when a user makes a query.
- Uses search techniques (keywords, similarity matching, metadata filtering) to locate the best data.
- Prioritises accurate and useful results to provide the AI with solid reference material.

#### Generator – The Communicator

- Combines retrieved data with AI's language capabilities to create clear, well-structured responses.
- Ensures AI doesn't just pull facts but delivers them in a way that makes sense to the user.
- Balances AI creativity with factual accuracy, making responses more useful and reliable.

## RAG-as-a-Service

For companies looking to build their own RAG applications, Ragie<sup>19</sup> provides an effortless way to set up and manage RAG pipelines. By staying at the forefront of RAG optimisation, Ragie ensures its clients benefit from the latest advancements without the complexity of in-house development.

A recent case study highlights how Ellis, a legal-tech firm, used Ragie to automate legal brief drafting, integrating AI retrieval with their Google Drive documents, yielding a 5–10x speed increase in document preparation—without compromising accuracy or human oversight.<sup>20</sup>

## Advanced RAG for Enterprise Search

RAG is also increasingly incorporated in solutions with a broader set of capabilities, for specialised use cases. We picked out Akooda as one to watch in 2025. Akooda specialises in enterprise search, pushing RAG beyond document retrieval into real-time knowledge discovery. Using an advanced form of RAG built on knowledge graphs, Akooda connects people, projects, and insights across fragmented enterprise systems.

By going beyond static documents to map relationships between experts, workflows, and collaboration networks, Akooda provides an example of how RAG is evolving—ensuring enterprises access not just data, but the right people and context behind it.



	<b>Founded</b> 2020
	<b>HQ</b> Massachusetts US
	<b>Employees</b> 25+
	<b>Funding</b> \$11m, Seed (latest)
	<b>Category</b> Enterprise Search
	<b>Customers</b> Industry Agnostic

The modern enterprise struggles to break down silos and make sense of its scattered knowledge. Akooda is an AI-powered enterprise search solution, connecting people, projects, and insights in real time across all company tools.

## The Business Case

### The Problem

Modern enterprises face a growing challenge: information is scattered across numerous tools, platforms, and silos, making it increasingly difficult to access and leverage effectively. Employees spend hours each week searching for documents, expertise, or data insights, leading to inefficiencies and missed opportunities. Without a holistic view of organisational knowledge, companies struggle to optimise operations and unlock their full potential.

### Internal Solutions

Developing an in-house enterprise search platform to address these challenges is a monumental task. Organisations often underestimate the complexity involved in creating a system capable of understanding nuanced natural language queries and mapping relationships between disparate datasets. Security and privacy concerns further complicate internal efforts, requiring expertise in areas such as cybersecurity and permissions management.

### Akooda's Solution

Akooda transforms enterprise search by enabling organisations to fully leverage its knowledge assets. With AI-powered, real-time processing and a privacy-first architecture, Akooda eliminates traditional barriers to information discovery. It doesn't just locate documents—it maps relationships between people, projects, and insights, creating lean, actionable knowledge graphs across Slack, JIRA, Google Drive, and dozens of other tools. Akooda has allowed enterprises to save at least 9 hours per week per employee, who each make 10–15 high-value queries per day.

## What's Unique About Akooda?



Akooda differentiates itself with its privacy-first design and real-time enterprise search capabilities. Unlike many traditional solutions, it avoids storing raw customer data or relying on indexing, which makes it particularly well-suited for industries with stringent privacy and security needs, such as finance and healthcare.

The team's background in cybersecurity and cloud technology has informed a platform that focuses heavily on data security and compliance. Akooda's approach goes beyond simple document search by dynamically mapping relationships between people, topics, and files, offering a more contextual and connected view of organisational knowledge.

## AI Capabilities

### Natural Language Search

Akooda leverages advanced natural language understanding to answer complex business questions directly, such as "Who knows the most about X?" or "What's the status of this project?" Its ability to interpret nuanced queries allows organisations to find precise answers quickly, saving time and improving decision-making.

### Dynamic Knowledge Graphs

Akooda dynamically creates real-time knowledge graphs that map relationships between people, projects, and topics across platforms like Slack, JIRA, and Google Drive. This feature enables users to identify experts, understand collaboration networks, and access contextual information effortlessly, breaking down organisational silos.

### Controlling AI Risk

#### Privacy-First Architecture

Akooda's platform is designed to meet the highest standards of enterprise data privacy. It operates without storing raw customer data and inherits permissions from the underlying platforms, ensuring that only authorised users access sensitive information. Deleted files are immediately excluded from search results, providing unparalleled data control.

#### Evidence-Based Responses

Akooda eliminates the risk of hallucinated content by providing answers derived strictly from source material. All responses are tied directly to their original documents or interactions, ensuring accuracy and trustworthiness for business-critical decisions.

#### Proactive Insights

Akooda surfaces critical trends and insights before they become problems. By analysing data from diverse channels, it flags key issues such as unaddressed project blockers, under-discussed objectives, or significant customer concerns. Insights are presented in visual dashboards, and help users prioritise evolving projects, KPIs, and customer needs.

#### Personalization

Akooda uses advanced deep learning models to tailor search results and insights to individual preferences and organisational terminology. By adapting to internal corporate language and user behaviours, Akooda delivers highly relevant, actionable information that aligns with each user's workflow.

#### No Raw Data Indexing Policy

Unlike traditional enterprise search tools, Akooda operates in real time without indexing or duplicating raw data. This approach minimises security risks and ensures that the company never takes ownership of any customer data, maintaining full compliance with enterprise security protocols.

#### Human Feedback Loop

Akooda's results are continuously refined through a human-in-the-loop feedback mechanism. Users can rate results, flag inaccuracies, and provide insights, enabling the system to improve dynamically while ensuring that human expertise remains central to the decision-making process.

## Ensemble Models

### What You Need To Know: What are Ensemble Models?

Ensemble models refer to AI systems that **combine multiple models to improve overall performance**, accuracy, and reliability. These models can work in different ways:

- **Model Blending:**

Multiple models generate predictions, and their outputs are combined (e.g., averaging or weighted voting).

- **Mixture of Experts (MoE):**

Different models specialise in different tasks, and a gating mechanism determines which model is best suited for each input.

- **Multi-Stage Pipelines:**

Different models handle different subtasks within a workflow, passing outputs between them for more refined results.

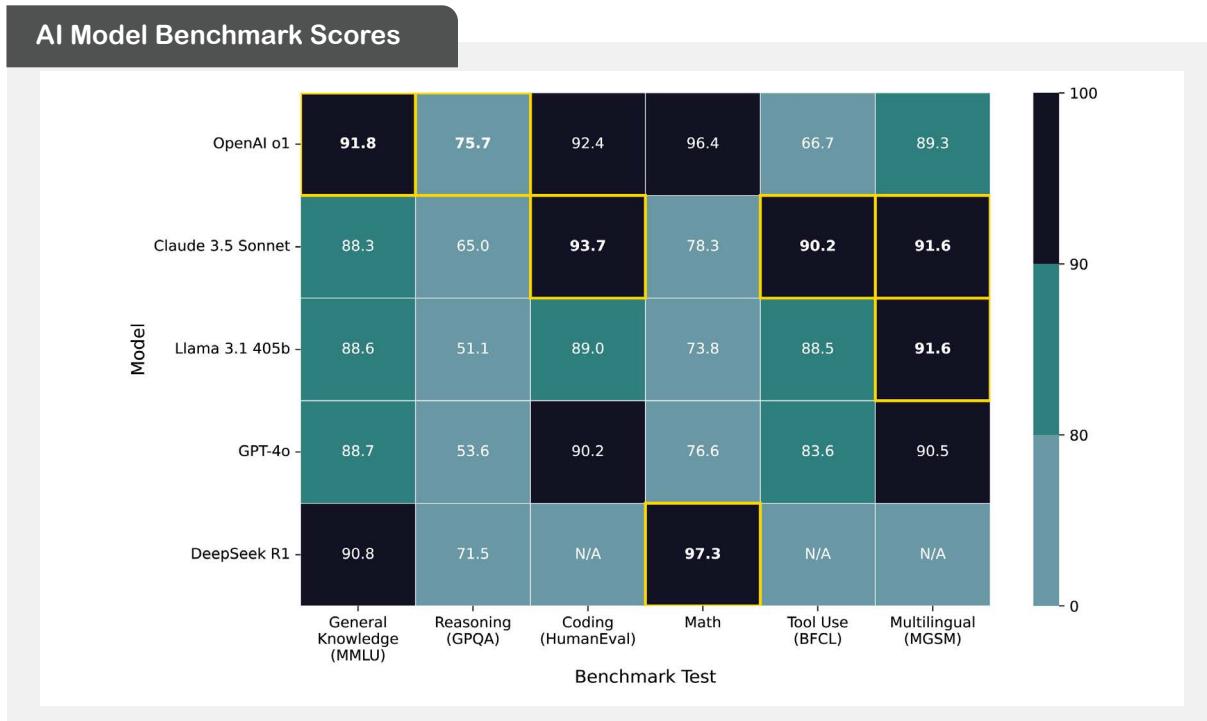
For example:

- One model might **extract key information** from a legal document.
- Another might summarise findings **for quick review**.
- A third might **fact-check** against an internal database to ensure accuracy.

This approach allows AI systems to deliver higher-quality results by **leveraging the best capabilities of each model**.

### What You Should Know: The Landscape of Capabilities

Ensemble models are gaining popularity as organisations seek **more adaptable, task optimised AI systems**. Different models excel in different areas:



Different models have different biases, training sets, and reasoning styles—by combining them, ensemble models compensate for individual weaknesses. **A carefully weighted combination ensures balanced, high-quality outputs.**

## What Experts are Talking About: When to Use Ensemble Models?

Ensemble models are particularly valuable in situations where a single AI model struggles with accuracy or where **combining multiple perspectives leads to better outputs**. Scenarios where ensemble models may fit:

- **Low-Accuracy Tasks:**

Certain tasks, like forecasting future events, are inherently difficult for a single AI model to predict accurately. A 2024 study by Schoenegger et al. found that an **ensemble of 12 diverse LLMs matched human crowd accuracy** in forecasting<sup>21</sup>, demonstrating how aggregating multiple independent models smooths out errors and improves reliability.

- **Reducing Systematic Biases in AI:**

Individual AI models can exhibit systematic biases, such as acquiescence bias (tendency to agree with statements). Ensemble models mitigate these biases by combining models trained on different datasets, ensuring that **no single bias dominates** the output, resulting in more balanced and neutral responses.

- **Uncertainty Estimation & Risk Reduction:**

Ensembles allow AI systems to quantify confidence levels by analysing model agreement. If all models converge on a response, confidence is high; if they disagree significantly, it **signals that human review may be necessary**—a critical feature in high-stakes applications like finance, legal, or healthcare AI.

The similarities between AI agents and ensemble models mean that the excitement and attention given to the orchestration of agents, particularly through MoE frameworks, also benefits ensemble approaches. Further, the increasing delivery of AI through APIs makes ensemble approaches even more natural.

## Using the Best Models for the Best Plaintiff Outcomes

Ensemble models are being applied in high-precision domains where accuracy and reliability are paramount. One standout example is Eve Legal, a LegalTech firm leveraging ensemble AI to streamline plaintiff-side legal workflows.

Eve uses a hybrid system of multiple LLMs, selecting the best model for each task based on complexity and domain specificity. This allows it to handle document automation, case synthesis, and damage calculations with greater accuracy than a single AI model could achieve.





Plaintiff law firms often face resource constraints that hinder their efficiency and growth. **Eve's AI-driven platform streamlines case workflows from intake to resolution, enabling firms to handle more cases with existing staff.**

	<b>Founded</b> 2016
	<b>HQ</b> California, US
	<b>Employees</b> 20+
	<b>Funding</b> \$47m, Series A
	<b>Category</b> LegalTech
	<b>Customers</b> Plaintiff Law Firms

## The Business Case

### The Problem

Plaintiff-side legal professionals face a fundamental challenge: the inability to scale efficiently. The sector is burdened by labour-intensive workflows that rely entirely on human effort. Tasks such as gathering evidence, drafting demand letters, managing document lifecycles, and navigating litigation involve significant manual effort and expertise. The inability to scale through technology limits their ability to deliver justice efficiently.

### Internal Solutions

Building an in-house solution to address these inefficiencies is a daunting prospect for most law firms. Developing high-quality AI requires a team of world-class engineering talent. Risks such as hallucination adds layers of complexity to AI implementation. Furthermore, internal development often lacks the iterative, focused design process needed to build a robust and effective technology solution.

### Eve's Solution

Eve simplifies legal workflows with unmatched ease—setup takes just one minute on a web browser. Unlike tools that address single tasks, Eve supports the full lifecycle of a case: uploading documents, synthesising case insights, and automating tasks like drafting demand letters and calculating damages. Combining advanced AI with a user-friendly design, its workflow-oriented approach enables firms to enhance their operational efficiency and serve more clients without expanding headcount.

## What's Unique About Eve?



Eve stands out for its AI-first approach to legal technology, a rarity in a space often dominated by solutions built by ex-lawyers. With a team of engineers from elite tech backgrounds (Google, Facebook, Microsoft), Eve takes a strategic and holistic approach to tackling legal workflows.

Unlike competitors, Eve focuses exclusively on plaintiff-side law, aligning its mission with empowering under-resourced legal teams. Its "price-per-case" model aligns financial incentives with customer success, while features like customisable playbooks, workflow blueprints, and damage calculations make it a comprehensive solution. Eve isn't just software—it's a movement to build an ecosystem of AI-native law firms.

## AI Capabilities

### Document Automation

Eve streamlines drafting with powerful document automation tools. Users can upload their own documents as "blueprints," allowing Eve to learn and replicate their preferred formatting, tone, and style. This enables it to produce customised outputs, such as demand letters or legal briefs, that fit seamlessly into a firm's workflow, significantly reducing the time needed for first drafts.

### Customisable Playbooks

Eve offers over 20 pre-built playbooks designed for common legal workflows, while also allowing users to create their own. These playbooks guide attorneys through complex cases step by step, ensuring Eve adapts to each firm's specific processes while maintaining flexibility for unique requirements.

### Controlling AI Risk

### Ensemble of LLM Models

Eve uses a hybrid system of multiple large language models (LLMs), including both closed and open-source models. Each task is directed to the most appropriate model based on its complexity and domain specificity.

### Risk-Based Validation:

Every reference or quoted material generated by Eve is cross-verified with the original user-uploaded documents. If the source cannot be validated, Eve flags it for review, adhering to a strict "trust but verify" standard that reduces the risk of hallucinated content.

### Evidence Synthesis and Case Insights

Eve excels at summarising the most important aspects of a case into concise, actionable case notes. Attorneys can ask specific, direct questions about uploaded documents and receive clear, relevant answers. Additionally, Eve flags potential "bad facts" and provides critical context to help legal teams assess risks and build stronger strategies.

### Automated Damage Calculations

Eve simplifies one of the most time-consuming aspects of case preparation by calculating economic and non-economic damages. Its AI is trained on relevant legal data, enabling it to produce accurate, case-specific figures, helping firms present well-supported claims while saving hours of manual effort.

### Strict Data Privacy

Eve ensures that case information remains fully segregated and secure. No data from one case is ever used in another, and Eve does not train its models on user-uploaded data.

### Human-Centric Design

Eve is built to support, not replace, human expertise, designed for attorney review and validation, with clear explanations provided for all recommendations. This human-in-the-loop approach ensures that final decisions remain with legal professionals.

## Custom Models

### What You Need to Know: What Are Custom Models?

Custom models are AI models **tailored to a specific organisation, industry, or task, allowing for deeper specialisation** beyond generic, off-the-shelf AI solutions. While customising AI once primarily meant fine-tuning an existing model with additional training data, today's approach increasingly involves training models from the ground up or significantly modifying base architectures to align with unique business needs.

Custom models are built with an organisation's proprietary data, industry knowledge, and domain-specific constraints—leading to more relevant, accurate, and controllable AI outputs.

### What You Should Know: Why Are Custom Models Gaining Popularity?

Customisation in AI isn't new—one survey found that at least 32.5% of businesses are already using fine-tuned models<sup>22</sup>, and another reported that 64% are developing custom applications<sup>23</sup> (though not necessarily custom models)—but it's becoming significantly more accessible. Adaptive ML<sup>24</sup>, for example, now offers **AI training as a managed service**, which includes personalised evaluations and deployment with a proprietary inference engine designed to combine convenience, safety and cost-efficiency. (See the [next section](#) for more details.)

Beyond specialisation and performance, **ownership and control** are driving factors behind the rise of custom models. Businesses are increasingly prioritising:

- **Independence from third-party providers:**

Avoiding an overreliance on external models, including vendor concentration risks.

- **Resource efficiency:**

Eliminating high markups from commercial providers and **avoiding the irony of using large language models for a small scope of tasks**.

- **Data security & governance:**

Maintaining full control over proprietary data rather than exposing sensitive information to external AI providers.

Overall, custom models offer:

- **Higher accuracy** – Optimised for **domain-specific language, regulations, and workflows**.

- **Better alignment** – Built to reflect **company policies, brand voice, and proprietary knowledge**.

- **More control** – Full **ownership and governance**, reducing third-party risks

### What Experts Are Talking About: Smarter, Smaller, More Efficient AI.

As AI development shifts toward specialised, cost-effective solutions, **models are shrinking**. Just as smartphones condensed powerful computing into pocket-sized devices, AI is transitioning from massive, general-purpose models to smaller, highly-optimised language models that deliver **better results with fewer resources**.

Here are three companies leading this shift, each with their own approach to building and utilising Small Language Models (SLMs):

- **Adaptive ML – Reinforcement Learning for Smarter Generalisation**

- **Malted AI – Knowledge Distillation for Efficiency**

- **Arcee – Small Agents for Big Businesses**



## Adaptive ML – Reinforcement Learning for Smarter Generalisation

Adaptive ML<sup>25</sup> enhances enterprise AI by evaluating, tuning, and deploying language models. The platform offers features such as one-click tuning, efficient learning through synthetic data generation, and continuous performance refinement based on user feedback and business metrics. Crucially, Adaptive ML focuses on using reinforcement learning (RL) to build small, specialised models that outperform popular frontier alternatives.

### The Power of Reinforcement Learning

Much of AI fine-tuning today relies on Supervised Fine-Tuning (SFT)—where models are trained using human-labeled datasets to ensure specific behaviours. While effective for instruction following and censoring harmful outputs, SFT has a critical limitation: it struggles to generalise beyond its training data.

A 2025 study involving Google DeepMind and UC Berkeley<sup>26</sup>, "SFT Memorises, RL Generalises," found that Reinforcement Learning (RL) consistently improves generalisation across tasks, while SFT severely degrades performance on out-of-domain (OOD) evaluations.

- RL boosted OOD success rates by **+3.5% to +61.1%**, depending on the task.
- SFT, in contrast, reduced OOD performance by **-5.6% to -79.5%**, showing it forces models to memorise rather than generalise.

DeepSeek's recent breakthrough with their R1 model underscores the potential of RL in AI development. By minimising human intervention and allowing the model to train itself through RL, DeepSeek achieved significant cost reductions and performance improvements. This success has provided additional credibility and renewed interest in RL as a key approach to achieving AI breakthroughs efficiently.

## Malted AI – Knowledge Distillation for Efficiency

Malted AI<sup>27</sup> specialises in knowledge distillation, a process that transfers knowledge from a large, complex AI model (the "teacher") to a smaller, more efficient model (the "student"). This enables the student model to retain the power of its larger counterpart while being significantly smaller, faster, and cheaper to run.

By generating high quality synthetic data, Malted not only packages but enhances the distillation process, allowing for the creation of Small Language Models (SLMs) that are 10-100 times smaller than traditional LLMs.

### The Power of Knowledge Distillation

Knowledge distillation involves training a student model to replicate the behaviour of a teacher model by learning from the teacher's outputs. These outputs are termed "soft labels," unlike hard labels (which provide definitive answers), soft labels offer probability distributions over possible outcomes, capturing the teacher's confidence and understanding of relationships between concepts. This nuanced information enables the student model to generalise better and perform effectively even with limited data.

Malted's synthetic data capabilities power their distillation pipeline—an approach that is supported by independent research. Xu et al. (2024) published an in-depth study, "A Survey on Knowledge Distillation of Large Language Models", emphasising the critical role of data augmentation in enhancing knowledge distillation processes:

 By leveraging data augmentation to generate context-rich, skill-specific training data, knowledge distillation transcends traditional boundaries, enabling open-source models to approximate the contextual adeptness, ethical alignment, and deep semantic insights characteristic of their proprietary counterparts.



DeepSeek's recent breakthrough with their R1 model is a useful example, again. By employing distillation techniques, DeepSeek developed a competitive AI model at a fraction of the typical cost and computational resources. This achievement underscores the potential of distillation to democratise access to advanced AI capabilities, making them more accessible and affordable.

## Arcee – Small Agents for Big Businesses

Arcee AI<sup>28</sup> focuses on developing Small Language Models (SLMs) to provide efficient and specialised AI solutions for enterprises. Their offerings include:



### Out-of-the-box models:

Arcee offers a suite of purpose-built SLMs designed for specific tasks and data. These models are optimised for speed and cost-effectiveness, making them ideal for various applications. For instance, the "Virtuoso" series caters to general-purpose tasks, while the "Coder" series is tailored for programming and development needs.



### Model merging:

Arcee's MergeKit is an industry-leading tool that allows users to combine models into a single, more effective model without additional training. This process enhances capabilities, without increasing the scale or inference cost of the final model. It means that companies can train just one model on their data, and stretch that knowledge across more capabilities by merging with another model.



### Agents:

Arcee Orchestra is a specialised AI agent automation platform designed to optimise business workflows through Small Language Models (SLMs). Unlike generic AI solutions, Orchestra enables enterprises to deploy tailored AI agents for specific operational tasks, ensuring domain-relevant accuracy and efficiency. By orchestrating multiple SLMs—each finetuned for distinct functions—the platform enhances response quality, supports multi-step process automation, and reduces computational overhead.

Arcee's push into AI agents is well-timed, given the significant momentum that this architecture is bringing into 2025. See the next section for more details.

## Beyond Chatbots

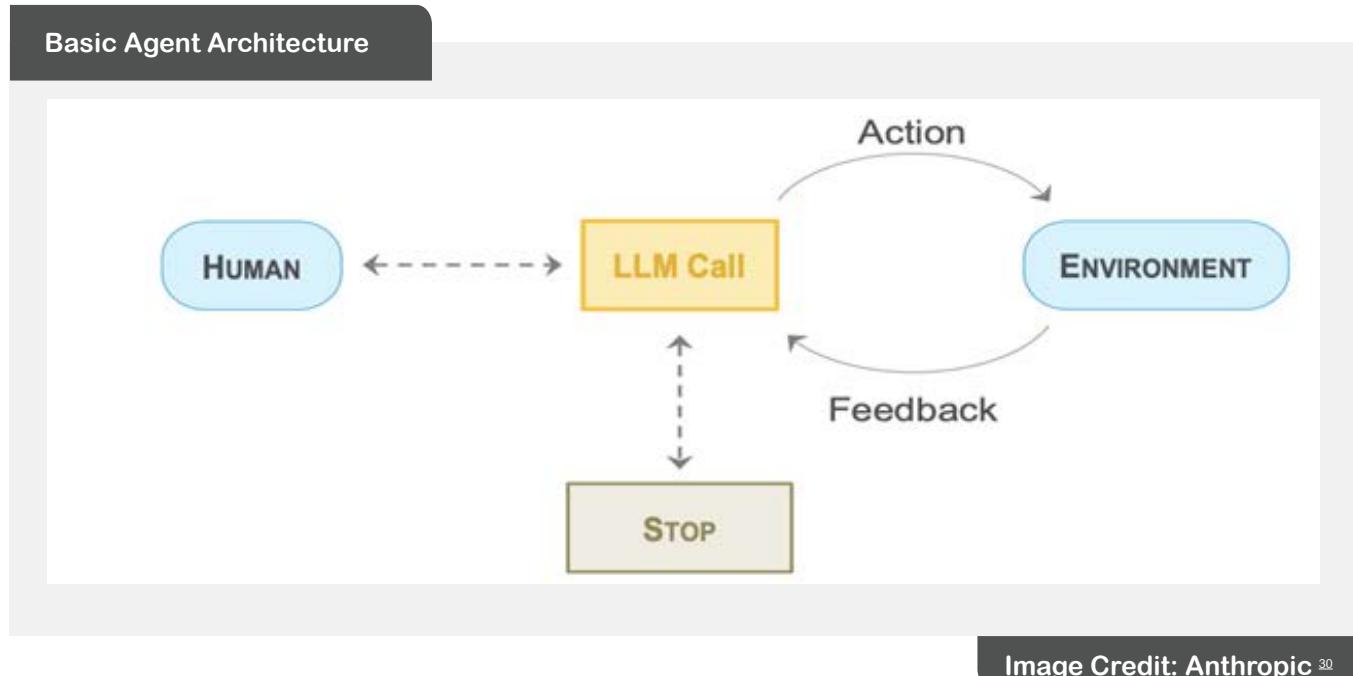
Chatbots have proven popular due to the intuitive nature of natural language queries and conversations. However, not all insights are optimally derived in a chat log. While chatbots are especially helpful when users know what they don't know—when they know what to ask—and when answers are best formed in text, other mechanisms may be preferred when this is not the case.

S&P Global forecasts<sup>29</sup> that text generators are set to experience the lowest CAGR (49.9%) of all GenAI segments and are set to lose the most market share from now until 2023-2028. Further, foundation models stand to lose ground in those same metrics. This underscores the shift toward highly practical and proactive forms of AI.



## Agents

### What You Need To Know: What Are Agents?



AI agents are configurations of AI models designed to operate autonomously, execute tasks, make decisions, and interact with systems or users.

#### Parker & Lawrence's Definition

While many definitions of AI agents resemble descriptions of chatbots like ChatGPT, we propose a minimalistic definition to distinguish true AI agents. AI agents should have three key characteristics:

- **Autonomy** – AI agents initiate and execute tasks without continuous human intervention, demonstrating the capacity to act independently rather than merely responding to explicit prompts.
- **Objectives** – AI agents operate according to a persistent, sustained objective, allowing them to make sequential, contextually relevant decisions over time. This is implied by autonomy—as agents need some way to act coherently in the absence of repeated prompting.
- **Access** – AI agents are not limited to pre-trained knowledge; they can retrieve and process information from external sources, including APIs, databases, and digital tools, enabling dynamic adaptation to new data and evolving environments. This includes access to their own past actions and interactions, in other words, they have memory.

For enterprises, this transition from passive AI models to proactive digital agents marks a fundamental shift in AI's role—moving beyond information retrieval to workflow automation, operational efficiency, and real-time decision-making with less human oversight.



## What You Should Know: The Growth Trajectory of Agents



Just as there is an app for everything, there will be an AI agent for everything.

— Ajit Tripathi, Hadron Founders Club



AI agents are rapidly becoming a focal point in AI strategies—BigTech references to agentic AI increased 17x in 2024—reflecting a significant shift towards automation and intelligent task management.

- ▶ **Investment:** Autonomous agents and digital co-workers saw the biggest growth in VC deal activity in 2024, at 150%.<sup>31</sup>
- ▶ **Early adoption:** LangChain's 2024 research found that approximately 51% of surveyed organisations have implemented AI agents in production environments, with mid-sized companies (100–2,000 employees) leading at 63% adoption, and 78% are actively developing agents.<sup>32</sup>
- ▶ **Enterprise growth:** Deloitte Global forecasts that 25% of enterprises that use GenAI will deploy agents in 2025, increasing to 50% by 2027,<sup>33</sup> while BCG's survey of enterprise executives finds that 67% are considering the use of agents, with almost a third viewing them as having a central role in their AI strategy.

In 2024, major cloud providers moved to capitalise on the growing momentum of agentic AI. AWS introduced Bedrock Agents, a managed service that simplifies the deployment of autonomous AI agents within enterprise applications.<sup>34</sup> Microsoft launched its AI Agent Service to automate tasks like customer support and workflow management,<sup>35</sup> while Oracle unveiled AI agents designed to streamline finance, sales, and supply chain operations.<sup>36</sup>

This progress is material, and is enough to warrant excitement, however, the current frenzy around agents is also driven by **bold predictions** from industry leaders:

### AI Agents Will Enter the Workforce



“In 2025, we may see the first AI agents ‘join the workforce’ and materially change the output of companies.”

— Sam Altman, CEO, OpenAI





"The IT department of every company is going to be the HR department of AI agents in the future... Today they manage and maintain a bunch of software from the IT industry; in the future they will maintain, nurture, onboard, and improve a whole bunch of digital agents."

— Jensen Huang, Founder & CEO, NVIDIA



"AI agents will become the primary way we interact with computers in the future. They will be able to understand our needs and preferences and proactively help us with tasks and decision-making."

— Satya Nadella, CEO, Microsoft

## AI Agents Will Reshape the Economy



"Agentic AI will power a startup golden age. Cloud computing moved the economics of entrepreneurship from CAPEX to OPEX. Agentic AI moves them from employee payroll to software subscription. Digital banks and regulated FinTechs will be able to leverage their tech infrastructure and licenses to grow even faster than before."

— Huy Nguyen Trieu, Co-founder, Centre for Finance, Technology and Entrepreneurship (CFTE)



"We think the hottest new area of venture investment will be autonomous AI agents—software agents that can simulate human behaviour and plan, make decisions, and execute tasks in complex environments without human intervention or supervision."

— Vibhor Rastogi, Global Head of AI/ML Investments, Citi Ventures





"A future world of AI agents, combined with IoT (Internet of Things), may need to make autonomous micropayments for data or energy it uses as part of their ongoing activity."

– Citi Research



## AI Agents Will Accelerate the AI Revolution



"Agents are not only going to change how everyone interacts with computers, they're also going to upend the software industry, bringing about the biggest revolution in computing."

– Bill Gates, Co-founder, Microsoft



"We've seen that with the great technological revolutions of the past. Each technological revolution has gotten faster, and this will be the fastest by far. Helpful agents are poised to become AI's killer function."

– Sam Altman, CEO, OpenAI



"AGI will take the form of some kind of an AI agent."

– Andrej Karpathy, Former Director of AI at Tesla



## What Experts Are Talking About: The Path to Real AI Agents

AI agents are already delivering measurable productivity gains in certain use cases.

A study using data from 5,179 customer support agents found that AI-powered conversational assistants increased productivity by 14% on average, with the most significant impact on novice and lower-skilled workers (a 34% improvement), while highly experienced employees saw minimal gains.<sup>37</sup>

Similarly, research from MIT, Princeton, and the University of Pennsylvania analysed **AI-assisted coding** across 4,800 developers<sup>38</sup> at Microsoft, Accenture, and a Fortune 100 company. Developers using GitHub Copilot saw:

- ▶ **26%** more tasks completed on average.
- ▶ **13.5%** increase in weekly code commits.
- ▶ **38.4%** rise in code compilation frequency.
- ▶ No observed decline in code quality.
- ▶ The biggest gains among junior developers, helping them bridge skill gaps faster.

Consumers are also increasingly comfortable with AI agents. Salesforce found that **37% of consumers are open to AI agents** generating personalised and useful content for them, rising to 44% among Gen Z users.<sup>39</sup>

### What is driving progress in agents?

The rapid advancement of AI agents is driven by several key trends that are shaping their capabilities and real-world applications. Three major technical factors are influencing their evolution:

#### 1. Smarter Models: Breakthroughs in Reasoning

AI models are becoming increasingly capable of complex reasoning, allowing agents to perform multi-step problem-solving with greater accuracy and stability. Developers are taking different approaches to achieving higher quality reasoning:

##### Deliberative Alignment: OpenAI's o3 Model

OpenAI's o3 model enhances reasoning through deliberative alignment, where the AI internally evaluates multiple possible solutions before selecting the most optimal one. Instead of presenting every step of its reasoning to the user, o3 searches, refines, and ranks different solution paths internally, ensuring that its final response is well-calibrated. This method enables AI agents to handle intricate tasks—such as coding and logical reasoning—with improved reliability. By aligning its internal reasoning steps with its final outputs, o3 helps AI agents make more stable and strategic decisions, particularly in tasks requiring precise judgment.

##### Chain-of-Thought: DeepSeek's R1 Model

DeepSeek's R1 model, on the other hand, adopts chain-of-thought (CoT) reasoning, where the AI explicitly shows its step-by-step thought process. Unlike deliberative alignment, which keeps internal reasoning hidden, CoT displays intermediate steps to the user, allowing for greater transparency and interpretability. By breaking down problems into sequential reasoning steps, DeepSeek's R1 enables user-auditable decision-making in complex domains.



## 2. Integrated Agents: Interfacing with External Systems

AI agents are increasingly integrating with external tools, databases, and applications, allowing them to access real-time data and perform actions beyond their initial training. This integration enables agents to execute tasks such as retrieving up-to-date information, interacting with software applications, and controlling IoT devices, thereby enhancing their functionality across various domains.

### Agents of Things

The evolution of the Internet of Things (IoT) is giving rise to the **Agents of Things (AoT)**—a paradigm where interconnected devices are enhanced with autonomous decision-making capabilities. This shift addresses previous IoT limitations.<sup>40</sup>

- ▶ While IoT enables devices to connect and exchange data, individual devices lack intelligence and decision-making capabilities.
- ▶ IoT relies on predefined rules and central processing rather than allowing devices to act autonomously.
- ▶ The sheer volume of connected devices leads to scalability challenges, security concerns, and inefficient decision-making.

The emergence of **world foundational models**<sup>41</sup> is driving this shift, providing a shared intelligence layer that enables AI agents to interact meaningfully with the physical world. These models are trained on multimodal datasets—incorporating sensor data, real-time video, spatial mapping, human interaction patterns, and industrial process logs—allowing them to understand and respond to their environments dynamically.



"The ChatGPT moment for physical AI and robotics is around the corner"

— Deepu Talla, Nvidia's vice-president of robotics

This shift is particularly transformative in robotics and autonomous systems. The global robotics market, projected to reach \$154 billion by 2033<sup>42</sup>, is being driven by AI-powered automation.

Companies like NVIDIA are advancing this space with investments in others, including Figure AI's \$675 million Series B in February 2024<sup>43</sup>, and hardware of its own, like Jetson Thor—designed to enhance the intelligence of humanoid robots and industrial automation systems<sup>44</sup>.

The World Economic Forum's January 2025 white paper<sup>45</sup>, "Frontier Technologies in Industrial Operations: The Rise of Artificial Intelligence Agents," explores how AI agents are transforming industrial operations. It highlights two primary types of AI agents:



- ▶ **Virtual AI Agents:** These software-based agents operate in digital environments, autonomously managing tasks such as process control and production planning.
- ▶ **Embodied AI Agents:** Integrated into physical systems like robots, these agents enable machines to perceive and interact with their surroundings, enhancing automation capabilities.

The report emphasises that adopting AI agents can lead to near-autonomous operations, improving efficiency and redefining human roles from manual operators to strategic overseers.

## Composable: Teaching Machines for Intelligent Industrial Automation

San Francisco-based startup **Composable**, founded by ex-Microsoft engineers and led by CEO **Kence Anderson**, is well-positioned to catch this wave. Their platform enables engineers to directly teach AI agents to operate autonomously across various equipment, including CNC machines, bulldozers, drones, robotic arms, and chillers.

Utilising a Machine Teaching methodology, Composable leverages operator expertise to break down complex tasks into individual skills, facilitating rapid training and deployment of intelligent agents. This collaborative approach between human expertise and AI-driven automation sets Composable apart in the industry<sup>46</sup>.

### 3. More Agents: Multi-Agent Systems

Multi-Agent Systems (MAS) enable multiple autonomous agents to collaborate, share information, and achieve complex goals beyond the capability of a single agent. These systems operate with autonomy, coordination, and scalability, allowing agents to dynamically adapt to their environment and optimise decision-making.

A MAS works by having each agent perceive its surroundings, communicate with others, and take independent or cooperative actions. Agents can assume specialised roles within a team, such as planners, executors, or communicators, to efficiently tackle complex tasks. This role differentiation allows agents to leverage their unique capabilities and collaborate effectively. SLMs are powering this trend by enabling the efficient deployment of specialised agents, making MAS more scalable.

### The Shift to Vertical AI

Enterprises are increasingly seeking a return on investment from AI initiatives. While horizontal solutions like ChatGPT have demonstrated impressive capabilities, their impact can be challenging to measure due to a lack of specialisation. The conversation has turned to vertical AI solutions, exemplified by highly specialised agents, as the next frontier for more tangible benefits and clearer value propositions.

### Bridging Agents and Data Analytics

Numbers Station provides AI agents to power enterprise data workflows. By deploying a system of specialised agents, Numbers Station automates complex data analytics tasks—from cleaning and organising data to generating actionable insights—all through natural language interactions. This agentic approach reduces the reliance on technical teams, making enterprise data more accessible and actionable.

A key enabler of this flexibility is **AI delivered via API**. Rather than locking businesses into a predefined user interface, Numbers Station's AI agents can be seamlessly embedded into existing workflows and frontends. This modular, API-based architecture ensures that enterprises can integrate AI automation into their unique data ecosystems, rather than restructuring their processes to fit a single tool.



	<b>Numbers Station</b>
	<b>Founded</b> 2021
	<b>HQ</b> Washington, US
	<b>Employees</b> 15+
	<b>Funding</b> \$12.5m, Series A (latest)
	<b>Category</b> Conversational Analytics
	<b>Customers</b> Industry Agnostic

Enterprise data workflows are often fragmented, inefficient and manual. **Numbers Station** empowers users to automate analytics workflows via an ecosystem of specialist AI Agents, all in natural language.

## The Business Case

### The Problem

Enterprises today face a persistent challenge: unlocking value from their increasingly complex and fragmented data systems. While analytics tools have advanced, tasks such as cleaning, organising, and analysing data remain resource-intensive and time-consuming. Business teams often struggle to access actionable insights without relying on technical experts, creating bottlenecks that slow decision-making. These inefficiencies limit scalability and prevent organisations from fully capitalising on their data assets.

### Internal Solutions

Organisations often turn to automation to address inefficiencies in their data processes, aiming to streamline workflows and reduce the burden on data teams. However, these efforts are frequently hampered by the sheer scale and complexity of modern data ecosystems, and the technical debt within legacy systems. While artificial intelligence holds the potential to transform analytics workflows, building an in-house AI solution is far beyond the reach of most companies, who simply lack the resources; time, expertise and infrastructure.

### Numbers Station's Solution

Numbers Station provides a novel approach to tackling data workflow inefficiencies. Its conversational AI platform enables users to interact with their data intuitively, reducing the reliance on technical teams and making insights more accessible across an organisation. By automating key stages of the analytics lifecycle—such as cleaning and unifying data, organising it through a semantic knowledge layer, and generating actionable visualisations—Numbers Station helps businesses streamline operations and focus on strategic decision-making.

## What's Unique About Numbers Station?

Numbers Station distinguishes itself with its AI-first approach to structured data analysis, in contrast to traditional analytics tools that often add AI as an afterthought. Founded by a team of AI and database experts from Stanford, the company blends cutting-edge technical innovation with a deep understanding of enterprise data challenges.

Numbers Station focuses on building modular, agentic workflows that adapt to diverse business needs, rather than relying on a one-size-fits-all solution. Its unified knowledge layer creates a rich, interconnected view of data that not only powers advanced analytics but also ensures transparency and traceability. Additionally, Numbers Station emphasises seamless integration. By offering API-based embedding, clients can incorporate its capabilities directly into their existing systems and interfaces.



## AI Capabilities

### Conversational Data Queries:

Numbers Station enables users to interact with their data using natural language, allowing business leaders to ask direct questions and receive clear, actionable answers. This intuitive interface bridges the gap between technical data analysis and business decision-making, democratising data access.

### Agentic Automation:

Numbers Station employs an agentic approach to automation, deploying specialised agents to perform tasks such as generating visualisations or executing actions in third-party platforms like Slack and Google Sheets. This system ensures that insights are not just delivered but translated into meaningful, automated workflows tailored to the user's needs.

## Controlling AI Risk

### Validation Agent:

Numbers Station employs a validation agent that assigns confidence levels—low, medium, or high—to all outputs. This ensures that users can assess the reliability of results and focus on validating critical insights.

### Source-Based Analysis:

To reduce hallucination risks, Numbers Station prioritizes direct data extraction from source systems. This approach ensures that outputs are grounded in verified data rather than assumptions or inferred information.



### Unified Knowledge Layer:

Numbers Station's unified knowledge layer combines a knowledge graph with an AI-powered semantic layer to manage and contextualise data. It organises analytics assets—dashboards, datasets, metrics, and transformations—into a structured, queryable system that is both machine-readable and human-auditable.

- **Metadata-Enriched Nodes:** Nodes represent analytics artefacts, such as dashboards and datasets, enriched with metadata, natural language descriptions, and historical queries for added context.
- **Relationship Mapping:** Directed edges connect nodes, capturing relationships like dashboard-to-dataset links, table joins, and metric derivations, ensuring data lineage is transparent.
- **Actionable Structure:** The layer transforms raw metadata into a navigable network, making it easier for users to trace relationships, verify logic, and interact with their data effectively.



### Privacy and Security:

The platform is SOC 2 compliant, ensuring robust data privacy and security protocols. It does not use client data to train its models, maintaining strict separation and confidentiality across all workflows.



### Human-in-the-Loop:

Designed to complement human expertise, Numbers Station integrates validation steps into its workflows. Users are encouraged to review and confirm key insights, maintaining human oversight in critical decisions.

## Semantic Layers

### What You Need to Know: What Are Semantic Layers?

AI is increasingly used to organise information into structured relationships, making massive datasets easier to interpret. When explicit, these structures are often referred to as semantic layers.

When training AI models, connections between concepts often emerge implicitly—for example, a chatbot trained on vast amounts of text develops an internal sense of how different topics relate. However, these relationships remain locked inside the model, inaccessible and unstructured. Semantic layers are different: their explicit purpose is to define and expose relationships, making them transparent, interpretable, and usable across different applications.

At the core of these semantic layers are ontologies and knowledge graphs. An ontology defines concepts and their relationships, allowing AI to categorise and connect information dynamically. Unlike rigid classifications, AI-driven ontologies evolve over time, using machine learning to detect patterns and associations in data. Knowledge graphs, similarly, structure relationships between entities, enabling smarter, more context-aware applications.

This AI-powered structuring happens behind the scenes. Users don't interact with the semantic layer itself—they experience its outputs: a more relevant search result, a personalised recommendation, or a well-organised dashboard. In this way, AI operates invisibly, yet it fundamentally shapes how businesses understand and use data.

### What You Should Know - The Power of Semantic layers

The power of semantic layers is realised when they are yours—built on your data, your domain, and your unique concepts. Unlike generic AI models, which may not fully understand the nuances of your business, a well-structured semantic layer captures your terminology, relationships, and industry-specific knowledge, making AI-powered insights more relevant and valuable.

Without a semantic layer, systems—including AI—are limited to disconnected concepts, incomplete understanding, and potentially generic outputs. While traditional databases may be structured, they lack the optimisation needed to power higher-level insights. Alternatively, manually built structures—such as human-curated taxonomies—are slow to develop, prone to bias, and difficult to scale.

AI-driven ontologies and knowledge graphs solve these challenges by automatically discovering relationships in your data. For example, in e-commerce, a knowledge graph can link products based on shared attributes, customer behaviour, and contextual relationships, enabling better search results, smarter recommendations, and more insightful analytics.

These structures are often built using unsupervised learning, which detects patterns without requiring human input. This makes semantic layers:

- ▶ **Data-driven** – Built directly from the data, free from human bias.
- ▶ **Dynamic** – Continuously updated as new information emerges.
- ▶ **Structured** – Optimising and enriching structured data with clear, meaningful connections.



By leveraging AI to create and maintain semantic layers, businesses gain more accurate, scalable, and efficient knowledge organisation—leading to deeper insights, better decision-making, and AI that truly understands your world.

## What Experts Are Talking About: Higher Order Insights

Semantic layers take care of the complexity at the foundational level, allowing higher-order insights to emerge. By structuring granular data into an understandable and accessible format, they create a transparent framework that powers visualisation tools, AI-driven search, and intelligent assistants.

### AI-Ready Data

Downstream from semantic layers, other AI models benefit from access to highly structured, connected, and contextually relevant information. With a well-defined knowledge base, AI agents don't need to infer relationships blindly—they can draw directly from an explicit, trusted source. This ensures more accurate, consistent, and meaningful outputs, whether in chatbots, search engines, or decision-support systems.

For example, Illumex builds enterprise knowledge graphs that automatically map, label, and contextualise structured data<sup>47</sup>. By creating a fully documented, custom ontology based on an organisation's own data, Illumex ensures that AI agents can navigate this structure with precision—delivering responses that are context-aware, accurate, and free from hallucination.

With this structured foundation, AI agents' outputs become deterministic—fully transparent and explainable, and always drawn exclusively from your data. Unlike GenAI models that may fabricate information, Illumex ensures that every AI-generated response is anchored in verified, organisational knowledge, eliminating hallucinations entirely. This is critical for AI governance, where auditability and transparency are top priorities.

## Visual Insights - Unlocking AI-Powered Discovery for CPG Data

Not all insights are best derived from raw data or text-based interactions. A picture speaks a thousand words, and a well-structured dashboard can summarise a thousand customer interactions. By providing clear, organised relationships between concepts, semantic layers enable businesses to move beyond isolated data points and uncover patterns, trends, and anomalies visually, at scale.

AI-driven semantic layers create a transparent, structured foundation that can be continuously updated with live data and translated into visual formats through dashboards, search interfaces, and discovery tools. This provides businesses with AI-powered insights in a highly accessible, interactive format.

This is especially important as AI adoption grows. AI-generated insights are increasingly expected, yet only 34% of companies are actively reskilling employees to collaborate with AI tools.<sup>48</sup> For AI to deliver value at scale, insights must be intuitive and accessible—making visual discovery a promising avenue to democratise AI-driven decision-making.

One company leading this transformation is Harmonya, which is redefining how businesses interact with structured data through AI-powered, real-time visual discovery. By building dynamic, unsupervised taxonomies and knowledge graphs, Harmonya enables CPG brands to track trends, optimise strategies, and make sense of their data faster than ever before.





**Harmonya**

	<b>Founded</b> 2021
	<b>HQ</b> New York, US
	<b>Employees</b> 50+
	<b>Funding</b> Undisclosed, Series B
	<b>Category</b> Product Data Insights
	<b>Customers</b> CPG Retailers

The consumer packaged goods (CPG) industry has failed to modernise its use and understanding of product data. **Harmonya aligns, enriches, analyses and operationalises CPG data at scale, driving growth through insight.**

## The Business Case

### The Problem

CPG companies and retailers face significant challenges managing and utilising product data. Each retailer's unique datasets—often inconsistent in format and taxonomy—make it difficult for brands to achieve a unified view of their product performance. Traditional approaches to harmonising data from multiple sources are labour-intensive and prone to errors. Furthermore, brands struggle to access actionable insights on consumer preferences and emerging trends, hindering their ability to innovate, optimise marketing strategies, defend and gain market share.

### Internal Solutions

Building a robust in-house solution to address these challenges is an uphill battle for most organisations. Developing proprietary AI models requires specialised expertise in machine learning and natural language processing, as well as access to vast, high-quality datasets. Even when such resources are available, managing the volume and variability of data across product categories and retailers is a massive undertaking. Many firms lack the infrastructure to scale these solutions or adapt them quickly to evolving market demands.

### Harmonya's Solution

Harmonya transforms product data management and insights with its comprehensive platform. By leveraging custom-trained large language models (LLMs) and aggregating data from over 300 retail and review sources, Harmonya delivers enriched, harmonised datasets tailored to client needs. Harmonya empowers CPG companies to identify emerging trends, validate innovations, and optimise marketing strategies with unparalleled speed and precision.

## What's Unique About Harmonya:

Harmonya's origin story is rooted in solving real-world pain points for CPG brands and retailers. From day one, the team worked directly with companies like Coca-Cola and Unilever to understand the challenges of managing fragmented product data across diverse retailers. This hands-on approach shaped a product designed for immediate market fit. With a team of AI experts who've developed high-stakes systems for organisations like the CIA, Harmonya combines deep technical expertise with a focus on practical, industry-driven solutions.

At the core of Harmonya's innovation is its use of unsupervised learning to dynamically build taxonomies, associate concepts, and identify trends. Instead of relying on predefined attribute schemas, Harmonya's platform aggregates product data to uncover patterns and concepts naturally. This enables brands to identify emerging concepts, track nuanced consumer preferences, and adapt taxonomies to their specific needs, empowering clients to stay ahead even in rapidly changing markets.

## AI Capabilities

### Product Data Enrichment:

Harmonya's AI enhances item master lists by extracting concept-based attributes and tags from diverse retail and review sources. Its proprietary large language models (LLMs), trained specifically on product-specific language and consumer dialogue, allow for the discovery of emerging trends and consumer preferences without predefined categories.

### Automated Attribution Models:

Harmonya creates custom attribution models tailored to client-specific taxonomies and schemas. These models harmonise and standardise product data, enabling brands to achieve consistency and accuracy across disparate datasets.

### Real-Time Insights Platform:

Harmonya combines transactional sales data with enriched attribute data to deliver actionable insights through intuitive, interactive visualisations. Users can explore granular trends, track market share changes, and identify key drivers of consumer behaviour with precision.

### Flexible Knowledge Graphs:

Harmonya builds dynamic knowledge graphs to bridge gaps in product data, allowing clients to query their datasets like a search engine. This capability provides answers to complex questions such as "Why is this product trending in a specific region?" with speed and clarity.

### Evidence-Based Analysis:

Harmonya anchors its insights in verified product data sourced from over 300 retailers and review platforms. Emerging trends and associations are presented with supporting evidence, ensuring that insights are backed by objective, aggregated data rather than anecdotal signals.

## Controlling AI Risk

### Custom AI Models for Domain Precision:

Harmonya employs proprietary large language models (LLMs) designed specifically for product descriptions and consumer language. By focusing exclusively on product data, these models minimise the risk of generating irrelevant or inaccurate insights.

### Human-in-the-Loop Validation:

Harmonya's platform incorporates a partnership-driven approach where customers retain control over AI outputs. Users can validate and refine model predictions, re-triggering the system as needed. This iterative process ensures accuracy and reliability over time.

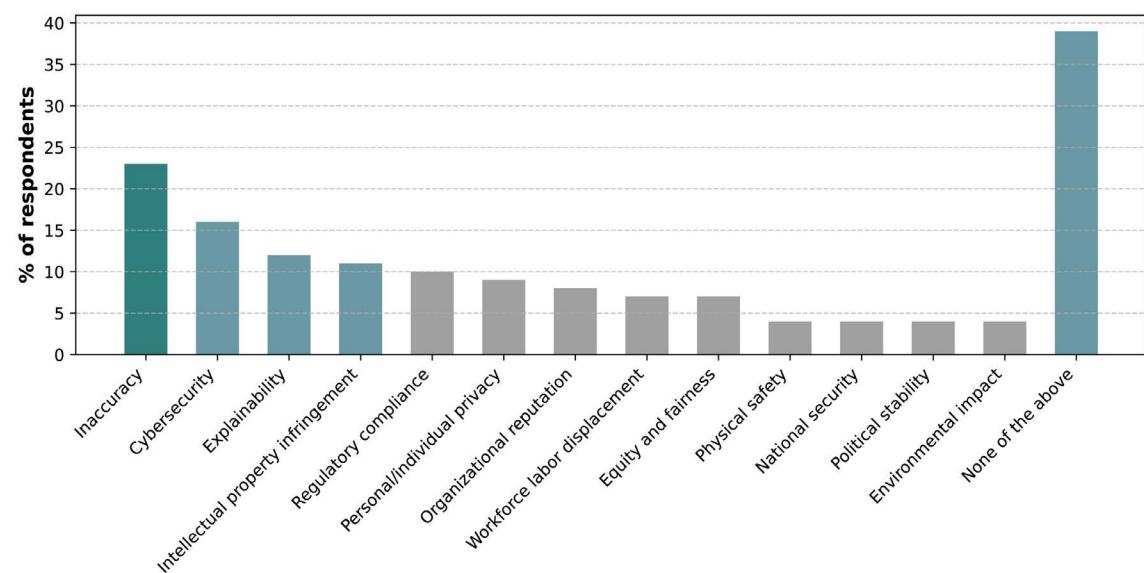
# Safety: Risk and Compliance

## Background

As AI becomes more deeply embedded in business and society, so too do the risks that come with it. From bias and misinformation to security vulnerabilities and regulatory uncertainty, the challenges surrounding AI safety are playing out in real time. Companies are already grappling with an expanding spectrum of risks, including data privacy concerns (66%), lack of control over AI decisions (48%), and regulatory compliance challenges (44%).<sup>49</sup> At the same time, the rapid rise of GenAI has introduced new threats, such as hallucinations, adversarial attacks, and intellectual property risks. These concerns have become so significant that **AI-related risk disclosures in Fortune 500 financial reports have increased by 473.5% since 2022.**<sup>50</sup>

It is surprising that so few large-scale AI-related incidents have occurred, especially since **more than 40% of developers claim not to perform evaluations on their AI.**<sup>51</sup> It may be that we are too early in the AI revolution to see such incidents, or that incidents are not being reported. McKinsey have found that negative consequences are manifesting for organisations:

GenAI Risks That Have Caused Negative Consequences for Organisations



Regulators around the world are responding, but their approaches vary. The EU has taken a prescriptive approach, classifying AI systems by risk levels and imposing strict compliance obligations, whereas the US favours sector-specific guidelines and/or voluntary compliance mechanisms. Meanwhile, jurisdictions like Singapore and the UK are shaping AI governance through principles-based approaches that emphasise risk management without stifling innovation.

Businesses must now navigate an increasingly fragmented regulatory environment, ensuring they comply with different—and sometimes conflicting—requirements across jurisdictions. For some, this uncertainty is slowing AI adoption. Even tech giants like Apple have had to rethink their AI rollout strategies, with Apple Intelligence currently facing delays in the EU due to compliance concerns under the Digital Markets Act. Unsurprisingly, **46% of executives report difficulty ensuring compliance in AI use cases.**<sup>52</sup>



With AI's risks becoming more complex, ensuring safety requires a structured approach to risk management, internal governance, and ongoing evaluation. This section explores three key aspects of AI safety: defining risk, which examines the core risks and emerging regulatory responses; measuring risk, which looks at methods for assessing AI model performance, security, and fairness; and addressing risk, which covers governance strategies, compliance tools, and industry best practices for mitigating AI-related threats.

## Defining Risk

### Key Risks and Regulations

#### What You Need To Know: What are the main risks associated with AI?

AI risks are vast and complex, encompassing ethical, operational, and security concerns that continue to evolve alongside advancements in AI itself. The fragmented nature of existing AI risk frameworks has made it difficult for organisations, policymakers, and researchers to systematically address, leading to inconsistencies in risk identification and mitigation efforts.

To tackle this challenge, the AI Risk Repository has developed a centralised, comprehensive, and continuously updated database that consolidates insights from over 56 AI risk frameworks, cataloguing more than 1,000 risks. Its Domain Taxonomy is a useful framework to categorise and organise AI risks:



### MIT's AI Risk Taxonomy



The technical security landscape of GenAI is particularly complex, with systems facing significant threats across several areas.

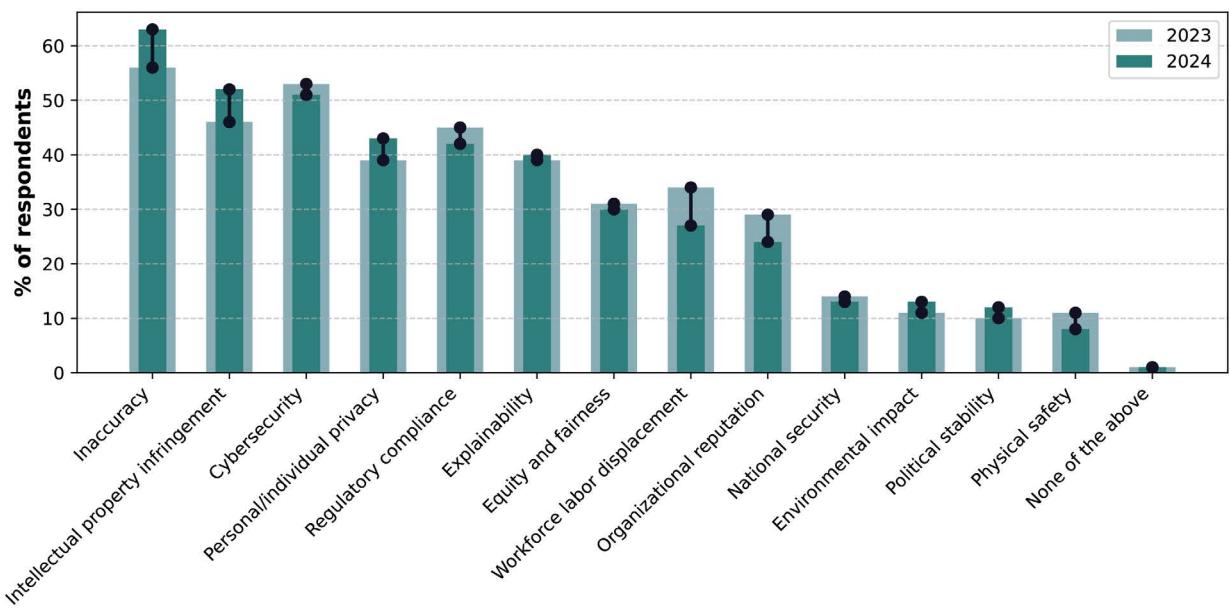
- ▶ Input manipulation vulnerabilities, such as prompt injection attacks and social engineering schemes, exploit user trust and expose sensitive information.
- ▶ Data exposure risks, including context and model leakage, compromise system confidentiality, while attackers may exfiltrate sensitive data during interactions.
- ▶ System integrity threats arise from jailbreaking attempts, RAG poisoning, and denial of service attacks targeting operational constraints.

Pillar Security's State of Attacks on GenAI report highlights the vulnerability of our current implementations:<sup>53</sup>

- ▶ **90%** of successful attacks resulted in sensitive data leakage.
- ▶ **20%** of jail break attack attempts successfully bypassed GenAI application guardrails.
- ▶ Adversaries require an average of **just 42 seconds to execute an attack.**
- ▶ Attackers needed only five interactions with the LLM per attack.

McKinsey's Global Survey on AI (N=1,363) found that organisations consider inaccuracy and intellectual property infringement as the most relevant risks associated with GenAI. Concern for both risks increased from 2023 to 2024:

#### GenAI Risks that Organisations Consider Relevant



## What You Should Know: Which Regulations Concern AI?

Regulators are aware of the risks that AI presents, and while some are hesitant to stifle innovation, many are responding with policies to protect against extreme events. The broad and still expanding footprint of AI means that it is captured under multiple regulatory domains:

### Model Risk

Model risk regulations ensure AI systems—particularly high-risk models—are reliable, explainable, and continuously monitored to prevent failures or biases.

### Key Regulations:

#### EU AI Act – European Union

- ▶ Introduces a risk-based framework requiring governance, transparency, and compliance measures for AI, particularly in high-risk applications.
- ▶ Mandates ongoing validation, explainability, and bias monitoring to prevent AI-related harm.
- ▶ Enforces post-market oversight, requiring firms to track AI performance and report failures to regulators.

#### The EU AI Act at a Glance<sup>54</sup>

Defining AI	Risk class	Examples	Requirements	TIMELINE AND IMPORTANT DEADLINES
"any machine-based system that operates autonomously to generate outputs such as predictions, recommendations, or decisions affecting physical or virtual environments."	Unacceptable Risk	Systems for targeted manipulation or suppression of human rights, such biometric categorization based on sensitive characteristics	Banned from the outset	 <b>2 Aug 2024:</b> Entry into law
<b>Scope</b>  The EU AI Act applies to all providers, deployers, importers, and distributors placing AI systems or general-purpose AI models on the EU market, regardless of whether they are located within the Union or in a third country, as well as to AI systems whose output is used in the EU.  <b>Providers:</b> Entities that develop or place AI systems on the EU market. <b>Deployers:</b> Organizations or individuals using AI systems in the EU. <b>Importers:</b> Entities bringing AI systems from outside the EU for sale or use within the EU. <b>Distributors:</b> Entities that supply or make AI systems available in the EU market without modifying them.	High Risk	Law enforcement, educational and vocational training, applicant assessment, credit assessment procedures, medical diagnostic systems	Risk management systems, technical documentation, human oversight, quality management systems, robust validation and cybersecurity	 <b>2 Feb 2025:</b> Ban on AI carrying unacceptable risk
	Limited Risk	Chatbots, deep fakes, personalized advertising	Transparency and disclosure requirements	 <b>2 May 2025:</b> Codes of Conduct applied
	Low Risk	Grammar checkers, spam filters	Voluntary Codes of Conduct	 <b>2 Aug 2025:</b> GPAI governance rules and obligations apply
<b>How will it be enforced?</b>  <b>Providers:</b> Providers initially categorize their AI systems based on the intended use and impact, conducting internal assessments to ensure compliance with the Act's criteria, particularly for high-risk systems. <b>Conformity Assessments &amp; Notified Bodies:</b> High-risk AI systems must undergo conformity assessments, either internally or through a notified body—an independent third-party organization designated by EU Member States. Notified bodies conduct thorough assessments to verify the system meets all regulatory standards before market entry. <b>Market Surveillance Authorities:</b> Once deployed, national authorities oversee compliance, with the power to investigate and intervene if risks or violations arise, ensuring the AI system remains safe and compliant. <b>Penalties:</b> Violations of the Act can result in fines up to €30 million or 6% of global annual turnover, depending on the severity of non-compliance.				 <b>2 Aug 2026:</b> Rules for high-risk systems apply
				 <b>2 Aug 2027:</b> Application of the entire EU AI Act for all risk categories



**SR 11-7 (US), SS1/23 (UK), and OSFI E-23 (Canada) – Model Risk Management (MRM) Frameworks**

- ▶ Originally designed for financial models but increasingly applied to AI-driven models.
- ▶ Require AI models to undergo conceptual soundness assessments before deployment.
- ▶ Emphasise explainability and independent validation to reduce black-box risks in AI-driven decision-making.
- ▶ AI models with significant business impact (e.g., credit scoring, risk assessment) must undergo stricter scrutiny.

## Data Privacy

Privacy regulations govern how AI systems process personal data, ensuring lawful, fair, and transparent data usage.

### Key Regulations:

**General Data Protection Regulation (GDPR) – European Union & United Kingdom**

- ▶ Imposes strict transparency requirements for AI-driven automated decision-making.
- ▶ Requires explainability, meaning businesses must disclose how AI models make decisions affecting individuals.
- ▶ Grants consumers the right to contest AI-driven decisions (e.g., credit approvals, hiring).
- ▶ AI systems must follow data minimisation principles, ensuring they only use necessary data.

**California Consumer Privacy Act (CCPA) – United States**

- ▶ AI-powered personalisation and profiling must be transparent to consumers.
- ▶ Grants individuals the right to opt out of AI-driven data processing for targeted advertising.
- ▶ Requires businesses to disclose whether AI is used for automated decision-making.

**Personal Information Protection and Electronic Documents Act (PIPEDA) – Canada**

- ▶ AI models must operate under explicit and informed consent when handling personal data.
- ▶ Requires AI-driven services to implement accountability measures for data governance.
- ▶ Introduces rules on automated decision-making, ensuring AI outputs can be explained to consumers.



**China's Personal Information Protection Law (PIPL) – China**

- ▶ Establishes strict rules on AI-driven facial recognition and biometric data usage.
- ▶ Requires firms to conduct risk assessments before deploying AI-based personal data processing.
- ▶ Mandates AI-generated recommendations (e.g., personalised ads) be transparent and explainable to users.

## Cybersecurity and Resilience

AI systems must be secure, resilient, and capable of withstanding cyber threats and operational disruptions.

### Key Regulations:

**Digital Operational Resilience Act (DORA) – European Union**

- ▶ AI models used in financial services must comply with stress testing and incident reporting requirements.
- ▶ Introduces mandatory risk management for AI-powered trading algorithms and fraud detection systems.
- ▶ Requires firms to document AI model security vulnerabilities and mitigation measures.

**Network and Information Systems Directive (NIS2) – European Union**

- ▶ Expands cybersecurity obligations to cover AI-based critical infrastructure.
- ▶ Requires AI models to be continuously monitored for cyber threats, such as adversarial attacks and data poisoning. Requires firms to document AI model security vulnerabilities and mitigation measures.
- ▶ Firms using AI in essential services (e.g., healthcare, finance) must implement real-time risk detection mechanisms.

**APRA CPS 230 – Australia**

- ▶ Effective from 1 July 2025, this standard focuses on operational risk management for financial institutions.
- ▶ AI Risk Management: Institutions using AI for financial services must identify, assess, and mitigate AI-related risks, such as bias, security vulnerabilities, and systemic failures.
- ▶ Third-Party AI Providers: Requires oversight of third-party risks, including AI vendors that supply critical financial services.



## Intellectual Property

Intellectual property regulations address the protection of creations of the mind, such as inventions, literary and artistic works, designs, and symbols, names, and images used in commerce, especially concerning AI-generated content.

### Key Considerations:

#### Copyright Laws – Global

- ▶ The legal status of AI-generated works varies by jurisdiction.
- ▶ In the United States, the U.S. Copyright Office has determined that works created without human involvement are not eligible for copyright protection.
- ▶ The United Kingdom is considering reforms to address AI and copyright, with debates focusing on balancing the interests of creators and technology developers.

#### Patent Laws – Global

- ▶ The question of whether AI can be recognised as an inventor has been tested in various jurisdictions.
- ▶ In the United States, the Federal Circuit ruled that only natural persons can be named as inventors on patent applications.
- ▶ The European Patent Office and the UK Intellectual Property Office have similarly held that an inventor must be a human being.

## The Convergence of AI and Technology Risk

AI's rapid expansion is blurring the lines between traditional risk categories, accelerating the need for a **unified approach: technology risk management (TRM)**. As AI becomes embedded in critical business operations—powering decisions, automating processes, and reshaping cybersecurity dynamics—organisations can no longer afford to manage AI-related risks in isolation.

Regulators are recognising this shift. The growing overlap between AI governance, cybersecurity frameworks, and operational resilience mandates indicates that **AI is not just an emerging model risk—it is now a core business risk**.

To keep pace, businesses need a **consolidated view of technology risk**.

**Riskonnect**, for example, provides organisations with an integrated technology risk management solution,<sup>55</sup> ensuring AI risks are proactively identified, monitored, and mitigated within a broader risk framework. Their capabilities include:

#### ▶ End-to-end visibility of AI and IT risks

Gain a single, centralised view of AI risks across IT, cybersecurity, and operational resilience. Track AI-related assets, threats, vulnerabilities, and regulatory compliance.

#### ▶ AI risk and control mapping

Link AI models to business-critical functions, mapping associated risks and controls to industry standards like DORA, NIS2, APRA CPS 230, and the EU AI Act.



**► Financial impact analysis**

Assess the potential financial consequences of AI failures, prioritising risk mitigation efforts based on real business impact.

**► Automated risk assessments**

Continuously monitor AI-driven risks, security gaps, and compliance status with built-in frameworks like NIST 800-53.

**► Proactive security and resilience measures**

Identify AI-related security vulnerabilities, evaluate control effectiveness, and integrate risk intelligence with cybersecurity programs.

## What Experts Are Talking About: Self-Regulation for Safety and Alignment

AI regulations are still taking shape, and many jurisdictions and industries have yet to establish comprehensive rules. In the meantime, leading AI companies are proactively developing internal frameworks to ensure the safe and ethical advancement of artificial intelligence. This self-regulation reflects the industry's recognition of AI's potential risks and its commitment to aligning technological progress with societal values—if powerful AI is misaligned, nobody wins.

### Anthropic's Responsible Scaling Policy

Anthropic has implemented a Responsible Scaling Policy (RSP) to maintain safety in AI development and deployment. The policy introduces AI Safety Level Standards (ASLs)—technical and operational safeguards that become stricter as models gain capability.

- **Establishes Capability Thresholds**, predefined risk checkpoints that trigger enhanced safeguards when crossed.
- **Requires higher ASL standards** for models with potential misuse risks, such as assisting in CBRN (Chemical, Biological, Radiological, and Nuclear) weapons development.
- **Enforces multi-stage risk assessments** to evaluate dangerous capabilities before deployment.
- **Mandates stronger governance controls**, such as appointing a Responsible Scaling Officer and maintaining whistleblower protections.

### DeepMind's Frontier Safety Framework

In 2024, DeepMind introduced their Frontier Safety Framework (FSF)—a structured protocol designed to mitigate severe risks associated with advanced AI capabilities.

- **Defines Critical Capability Levels (CCLs)**—specific capability thresholds at which AI models may pose heightened risks in areas such as autonomy, biosecurity, cybersecurity, and AI research acceleration.
- **Implements early warning evaluations** to detect when models approach CCLs, ensuring proactive risk management.
- **Applies security mitigations** to prevent model weight exfiltration, including access controls, high-trust developer environments, and confidential compute protections.
- **Enforces deployment mitigations** such as safety fine-tuning, misuse detection, and strict access controls to prevent the expression of critical capabilities.
- **Mandates response protocols** that may include pausing development or deployment if mitigations are insufficient.



## OpenAI's Preparedness Framework

OpenAI has introduced a Preparedness Framework, a structured system for evaluating AI risks at various capability levels.

- **Prioritises model evaluations** for risks such as cyber-offensive capabilities, deception risks, and autonomous replication.
- **Uses risk tiers** that trigger escalating safeguards as AI capabilities increase.
- **Incorporates rigorous internal testing**, third-party red-teaming, and access controls.
- **Engages independent auditors** for external oversight on alignment research and security measures.

## Frontier Model Forum: A Collective Industry Initiative

The Frontier Model Forum (FMF) was launched by Anthropic, Google DeepMind, Microsoft, and OpenAI to promote responsible AI development.

- Serves as a platform for **AI safety collaboration**, technical benchmarking, and risk mitigation strategies.
- Develops **shared evaluation frameworks** to assess AI security, societal risks, and reliability under stress.
- Engages with governments, academia, and civil society to **shape AI safety** policies and best practices.

## Safe Superintelligence Inc.

Ilya Sutskever, co-founder and former chief scientist of OpenAI, launched Safe Superintelligence Inc. (SSI) in 2024, aiming to develop safe superintelligent AI free from commercial pressures.

- Operates with a mission-first approach, **prioritising safety** over market competition.
- Secured over \$1 billion in funding in its initial months—showing **commercial appetite for risk averse initiatives**.
- Focuses on ensuring superintelligent **AI is aligned with human values** before widespread deployment.

# Measuring Risk

## Evaluations and Testing

### What You Need To Know: What are AI Evaluations?

AI evaluations are systematic methods used to assess model performance, reliability, and robustness. A comprehensive evaluation framework examines key elements of the AI system, including:

- **Inputs:** Evaluates the suitability and appropriateness of training and test data. This includes assessing data quality, representativeness, and potential biases that may affect model outcomes.
- .....
- **Model Architecture:** Examines the technical design and structure of the model, including its complexity, transparency, and scalability. Evaluation at this level helps ensure the model is both efficient and interpretable.
- .....
- **Outputs:** Assesses the accuracy, reliability, and consistency of model predictions, as well as any other outputs generated by the system. This includes testing for unintended consequences or errors that may arise in real-world use.



In particular, AI evaluations seek to assess:

➤ **Predictive Accuracy & Stability:**

Assessing how well the model performs on unseen data using metrics such as precision, recall, F1-score, and AUC-ROC. Stress testing for model drift ensures stable performance over time.

➤ **Bias & Fairness:**

Measuring whether the model produces disparate impacts across demographic groups. Fairness audits utilise statistical tests such as Demographic Parity, Equalised Odds, and the Disparate Impact Ratio.

➤ **Explainability & Interpretability:**

Ensuring that model decisions are transparent and understandable to stakeholders and regulators. Techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) help illuminate decision-making processes.

➤ **Robustness & Security:**

Identifying vulnerabilities to adversarial attacks, such as data poisoning or perturbation-based evasion. Evaluating model sensitivity to input variations and unexpected edge cases helps mitigate risks.

A well-structured AI evaluation framework not only minimises risks but also accelerates deployment and ensures regulatory compliance.

## What You Should Know: Model Validation is a Business Enabler

Despite the transformative potential of AI, most proof-of-concepts (PoCs) never reach production.<sup>56</sup> While concerns over accuracy, reliability, and bias are well-documented, the validation process itself often becomes an unintended barrier—particularly when conducted manually. Lengthy, fragmented, and inconsistent validation workflows slow time-to-market, erode stakeholder confidence, and increase regulatory and reputational risk.

Yet, AI validation isn't just a hurdle—it's a strategic enabler. Since all firms must validate their models, those that do it faster and better gain a competitive edge. A rigorous, structured validation framework doesn't just de-risk AI deployment; it generates insights that can be leveraged for future optimisation. Done well, it helps firms to:

➤ **Progress AI models** from PoC to production by systematically identifying and mitigating risks.

➤ **Optimise AI deployment** in line with risk appetite, ensuring high-quality validation provides an accurate picture of risk exposure.

➤ **Support compliance** with evolving AI regulations while enhancing trust and transparency in AI-driven decision-making.

<sup>56</sup> McKinsey (2020), said that about 87% of AI PoCs are not deployed in production. Generative AI has likely pushed these percentages even higher.



In an AI arms race, speed and efficiency are critical. The winners have already recognised that risk management is not a bottleneck to AI adoption—it is a vital step in the AI lifecycle, offering insights into both the model and the broader business case. However, deriving these insights efficiently and consistently is challenging, particularly with complex, opaque AI models. The [next section](#) explores what experts are doing to address this.

Model validation can therefore be a source of competitive advantage unlocking the fundamental transformations that two-thirds of digital and analytics leaders in banking believe GenAI will deliver. The potential rewards—an additional \$200 billion to \$340 billion, or 9-15% of operating profits annually—are so significant that another subplot has emerged: The race to validate AI is on.<sup>57</sup>

## What Experts Are Talking About: Automating and Expanding AI Evaluations

### The Need to Automate

Firms that treat AI validation as a continuous process—rather than a one-time hurdle—can deploy AI faster, mitigate risks more effectively, and drive measurable business impact with confidence. However, achieving this at **scale requires automation**.

The growing complexity and scale of AI models and their applications have made manual validation impractical. **Humans continue to play a critical role in risk management, but they can no longer play every role.** Instead, firms are increasingly using AI to validate and monitor AI—a shift that may sound futuristic but is already making an impact: **38% of firms currently use automated evaluation tools.**<sup>58</sup>

### Expanding to GenAI

The need for automation is especially pressing with GenAI, whose inputs, architecture, and outputs differ significantly from the traditional AI models firms are used to evaluating.

- **Inputs:** Unlike traditional AI, which processes static, structured data, GenAI handles freeform, multimodal inputs—text, images, audio, and video—within an iterative, conversational flow. Each prompt builds on previous exchanges, creating a dynamic feedback loop.
- .....
- **Model Architecture:** Classical models rely on predefined structures, while LLMs (transformers) operate at massive scale, incorporating context from entire interactions rather than just the latest input. Their emergent behaviours demand novel testing methods.
- .....
- **Outputs:** Traditional AI generates single, numerical predictions; GenAI produces evolving, open-ended responses. Evaluating performance means assessing coherence, factual consistency, and stability across multi-turn interactions rather than isolated outputs.

GenAI's iterative nature, contextual awareness, and non-deterministic outputs introduce new risks that traditional validation frameworks cannot fully address, requiring expanded, automation-driven testing strategies.<sup>59</sup>



## Validating GenAI

A comprehensive GenAI validation framework should evaluate models across three key dimensions:

### ► 1. Input-to-Representation (Embeddings & Retrieval)

Focuses on how well a model understands and encodes input data, including the quality of vector embeddings and retrieval mechanisms in knowledge-based AI.

**Key Tests:** Cosine similarity evaluations, retrieval efficiency benchmarking, sensitivity analysis.

### ► 2. Representation-to-Generation (Output Quality & Fidelity)

Evaluates the quality, coherence, and factual accuracy of generated outputs. Ensures AI-generated content aligns with expected outputs, minimising hallucinations.

**Key Tests:** BLEU, ROUGE, BERTScore, variance and diversity checks.

### ► 3. Performance and Impact (Security, Stability & Fairness)

Assesses the overall reliability, robustness, and ethical considerations of the AI system. This includes identifying vulnerabilities, including prompt injections and jailbreak attempts, and detecting disparities across demographics.

**Key Tests:** Red teaming, adversarial probing, fairness audits, explainability testing.

Validating GenAI applications requires a structured, automated approach—something experts are actively advancing. Platforms like **ValidMind** streamline this process by automating AI model testing (including GenAI), documentation, and validation. Their platform also structures workflows and collaboration, ensuring a more efficient and compliant model lifecycle.<sup>60</sup>

Purpose-built validation platforms help firms to:

- **Continuously test AI models** across different data distributions and conditions.
- **Embed validation directly into model pipelines**, ensuring consistency.
- **Rapidly iterate and improve models** through structured feedback loops and automated risk assessments.
- **Reduce validation timelines** from months to weeks—accelerating AI adoption.
- Improve **transparency and explainability** for regulatory compliance.

Organisations embedding AI risk reviews early in development cycles report higher compliance success rates. It is therefore no surprise that 58% of GenAI “high performers”—those attributing 10% or more of their EBIT to GenAI—already have testing and validation embedded in the release process of each model.



# Addressing Risk

## Governance and Tooling

### What You Need To Know: What is the role of Governance in Addressing AI Risk?

Governance is the mechanism that ensures accountability, oversight, and structured risk management in any organisation. As AI adoption scales, these same governance principles apply—but with new challenges unique to AI: models that evolve unpredictably, risks that emerge post-deployment, and regulatory frameworks that are still forming.

Leading AI adopters, having integrated AI deeper into their operations, face these risks most acutely.<sup>61</sup> Their experience shows that AI governance must adapt beyond traditional oversight models to remain effective.

#### How Governance Applies to AI:

##### Risk Management & Compliance → AI Risk & Regulatory Alignment

- Traditional governance ensures organisations comply with legal and operational risk standards. For AI, this means governing opaque, evolving models that require ongoing risk assessments, explainability measures, and compliance with shifting AI regulations.

##### Data Governance → AI-Specific Data Controls

- Standard governance mandates data integrity, privacy, and security. AI governance expands this by ensuring data is fit for model training, free from bias, and auditable, particularly for GenAI models that rely on dynamic, unstructured data sources.

##### Model Governance → Lifecycle Oversight & Validation

- While traditional governance ensures business processes are monitored and reviewed, AI governance demands continuous validation, post-deployment monitoring, and adversarial testing to mitigate risks like bias drift, hallucinations, and security vulnerabilities.

##### Product & Technology Governance → Responsible AI Deployment

- Corporate governance oversees product development to ensure regulatory and ethical standards are met. In AI, this requires embedding risk reviews at every stage—from model design to real-world deployment—ensuring AI-driven decisions are transparent, fair, and aligned with business objectives.

##### Operational Governance → AI Decision Accountability

- Traditional oversight assigns responsibility for financial, operational, and strategic outcomes. AI governance must do the same—establishing clear ownership for AI decisions, defining who is accountable when models fail, and ensuring human oversight remains in critical decision-making loops.



The complexity and speed of AI innovation puts increased emphasis on robust governance. These dynamics make the extent and eventualities of AI's risks difficult to comprehend for non-experts. Organisations cannot rely on their employees' common sense: less than 30% of organisations have provided GenAI training to more than 25% of their staff. And at present, risk management skills may be undervalued: only 33% of firms require AI teams to demonstrate risk mitigation expertise. **If humans cannot be expected to intuit risk, they must instead be given guardrails to manage it.**

Yet, despite its importance, **only 18% of organisations have an enterprise-wide AI governance framework**, leaving many exposed to unmanaged risks. Leading adopters, however, are addressing this gap; embedding legal oversight, enforcing early-stage AI risk reviews, and developing governance models that evolve alongside AI itself.

## What You Should Know: How Governance Tools are Unlocking AI for Enterprise

To implement AI governance at scale, enterprises are increasingly adopting specialised tools. Data governance, in particular, has become paramount due to the text-heavy and expansive nature of AI training data, the shift towards multi-modal systems encompassing even more data types, the maturity of data privacy regulations and the risk of biases in datasets. Further, BCG's survey of **1,800+ executives put data privacy and security at the top (66%) of AI risks** being navigated, and, of McKinsey's high performers, **70% reported difficulties defining data governance processes.**

Two solutions have caught our attention:

### ➤ Castlepoint: Visibility and Records Management

Castlepoint offers a metadata-driven approach to data governance, enabling organisations to gain complete visibility into their data assets. By automatically mapping and classifying data across the enterprise, Castlepoint helps identify sensitive, obsolete, or high-risk information, ensuring that AI systems access only appropriate and relevant data.

This process mitigates risks associated with AI hallucinations by preventing models from drawing on outdated or irrelevant data sources. Additionally, Castlepoint enhances auditability and regulatory compliance by maintaining detailed records of data usage and access patterns, which is crucial for organisations aiming to deploy AI responsibly.<sup>62</sup>



## ➤ Redactive: Access Control and Security

Redactive's Permissions Assurance platform identifies and resolves inappropriate data access at scale, preventing AI models and agents from exposing sensitive information due to overly permissive access controls. This enhances AI governance by:

- **Detecting Hidden Risks** – AI-native semantic analysis scans documents at a granular level, identifying inappropriate access to sensitive content that rule-based security systems often miss.
- **Continuous, Automated Protection** – Redactive automatically corrects misconfigured permissions in real time, ensuring that employees, applications, and AI models only access data they are explicitly authorised to use.
- **Preventing AI-Powered Data Leaks** – As AI exposes security gaps in existing data governance strategies, Redactive closes those gaps by mapping and controlling permissions at the document, sentence, and content-chunk level.<sup>63</sup>

## What Experts Are Talking About: Specialist Tools for Novel Risks

As AI risks evolve, **firms require specialised risk management tools** to counter emerging security threats unique to GenAI. Unlike traditional AI, GenAI introduces new attack surfaces and systemic vulnerabilities, requiring **AI-native security measures**. Key risks include:

- **Prompt Injections & Adversarial Attacks:** Manipulating model inputs to bypass safeguards, generate harmful content, or exfiltrate sensitive data.
- **Decision Integrity Threats:** Unintended model behaviours, data poisoning, or reinforcement of biased patterns that degrade trust in AI-driven decisions.
- **Cybersecurity Weaknesses in AI Pipelines:** AI-powered automation creates new entry points for model hijacking, supply chain attacks, and API vulnerabilities.

## Defending Against Emerging AI Threats

SplxAI is gaining traction for its AI-native security solutions, tackling unique GenAI security challenges. By deploying real-time threat detection and model integrity validation, SplxAI helps firms safeguard AI pipelines from manipulation and adversarial exploitation.





Generative AI and Large Language Models introduce new security risks that traditional defenses fail to catch. **SplxAI automates AI red teaming to detect and mitigate threats of LLM-powered apps and agents before and during runtime.**

	<b>Founded</b> 2023
	<b>HQ</b> New York, US
	<b>Employees</b> 20+
	<b>Funding</b> \$9m, Seed (latest)
	<b>Category</b> Cyber Security
	<b>Customers</b> Healthcare and Financial Services

## The Business Case

### The Problem

The rise of multi-modal AI systems and agentic workflows adds complexity to cyber security frameworks by increasing potential exploit channels. Traditional approaches lack specialized capabilities to detect sophisticated AI threats leaving at least ~30% of AI vulnerabilities undiscovered between testing cycles.

### Internal Solutions

Today, most organizations rely on manual red teaming conducted by internal security teams, which, while effective, demands significant time and expertise—up to three full-time engineers. These assessments often take weeks to complete, straining resources and creating bottlenecks that delay secure AI deployment.

### SplxAI's Solution

SplxAI automates AI security and red teaming, identifying and mitigating risks like toxicity, bias, hallucinations, and social engineering. It detects vulnerabilities such as prompt injections, context leakage, and data exfiltration—threats often missed by traditional security. By replacing manual red teaming with automated assessments, SplxAI cuts testing time from weeks to under an hour, executing 20+ predefined attack scenarios while allowing custom risk assessments via natural language. Seamlessly integrating into CI/CD pipelines, it ensures continuous AI validation at scale.

## What's Unique About SplxAI?

SplxAI is the first multi-modal AI security testing platform, covering text, images, voice, and documents for a comprehensive risk assessment.

Its customisable probes enable domain-specific tests, such as refining chatbot assessments to prevent harmful or misleading insurance advice. Each probe run generates detailed results, including simulated conversations that expose vulnerabilities through malicious prompts, helping users diagnose and resolve security gaps faster.

SplxAI also launched Agentic Radar, an open-source AI transparency tool for security engineers. It maps agent structures, identifies dependencies, and highlights vulnerabilities based on OWASP's Top 10 for LLMs and Agentic Threats. By visualising risks in multi-agent workflows, security teams can better understand weaknesses and tailor risk assessments to enhance AI defences.

## Controlling AI Risk

### Probe Platform:

End-to-end vulnerability assessment tool that automates and scales AI red teaming. It offers over 20 predefined probes like prompt injection, context leakage, fake news, and social engineering, each including multiple attack variations for detailed assessment. For example, the Jailbreak Probe assesses whether the AI can be manipulated to bypass constraints.

### Dynamic Remediation Strategies:

Offers tailored remediation steps, such as a built-in system prompt hardening tool, canary word detection for monitoring prompt leakage, and input/output filters for enhanced security. Users can track and execute these mitigation steps directly in tools like Jira or ServiceNow, streamlining the process.

### LLM Log Analysis:

A continuous monitoring system that processes JSON-formatted interaction logs from LLMs to identify potential security breaches and safety violations. When analysing multi-agent LLM workflows, the system can track threats across LLM-to-LLM interactions by examining the complete conversation chain. The platform performs automated triage by categorizing and prioritizing detected incidents based on their severity and success rate, providing detailed attack forensics including the exact queries that triggered security violations and explanations of why they were flagged as malicious. This enables near real-time threat detection and incident response for LLM deployments while eliminating the need for manual log review.

### Dynamic Threat Database:

SplxAI maintains a continuously updated threat database that includes emerging attack vectors and strategies gathered through frameworks like MITRE ATLAS, NIST, and OWASP Top 10 for LLMs, as well as proactive web scraping and cross-industry knowledge sharing. The database serves as the backbone for generating new test scenarios, helping users stay ahead of evolving threats and ensuring their AI systems are always protected against the latest attack techniques. SplxAI's proprietary taxonomy is another unique aspect, designed to capture and categorize emerging threats comprehensively.



# Scale: Infrastructure and Energy

## Background

AI is scaling rapidly, and infrastructure must keep pace. **83% of organisations** expect their AI workloads to grow within the next two years, while **two-thirds** anticipate needing significant upgrades to sustain this expansion.<sup>64</sup> Yet, scaling AI isn't just a matter of investment—it's constrained by energy, efficiency, and raw computing power.

Efficiency is becoming a bottleneck. **37% of companies** already cite the high cost of running AI models as a major challenge,<sup>65</sup> but the real problem is likely much bigger—many aren't even at the scaling stage yet. More critically, the **limiting factor isn't just cost, but physical constraints:** the U.S. may need to **double its power grid capacity** over the next decade just to keep up with AI-driven demand.

At the same time, a paradigm shift is emerging. DeepSeek's breakthrough in efficiency has reshaped expectations for AI performance at scale, while hyperscalers like Google and Microsoft continue to invest tens of billions into cloud capacity to meet surging demand (In mid-2024, Google Cloud's revenue had surged 28% year-over-year, more than double its parent company's overall growth rate.)<sup>66</sup>

Specialised AI chips present another avenue to improve AI's efficiency. Tech firms are pouring billions into AI chip investments—ByteDance alone plans to spend \$12 billion on AI infrastructure in 2025, including **\$5.5 billion on AI chips** in China and \$6.8 billion overseas to expand its foundation model training.<sup>67</sup>

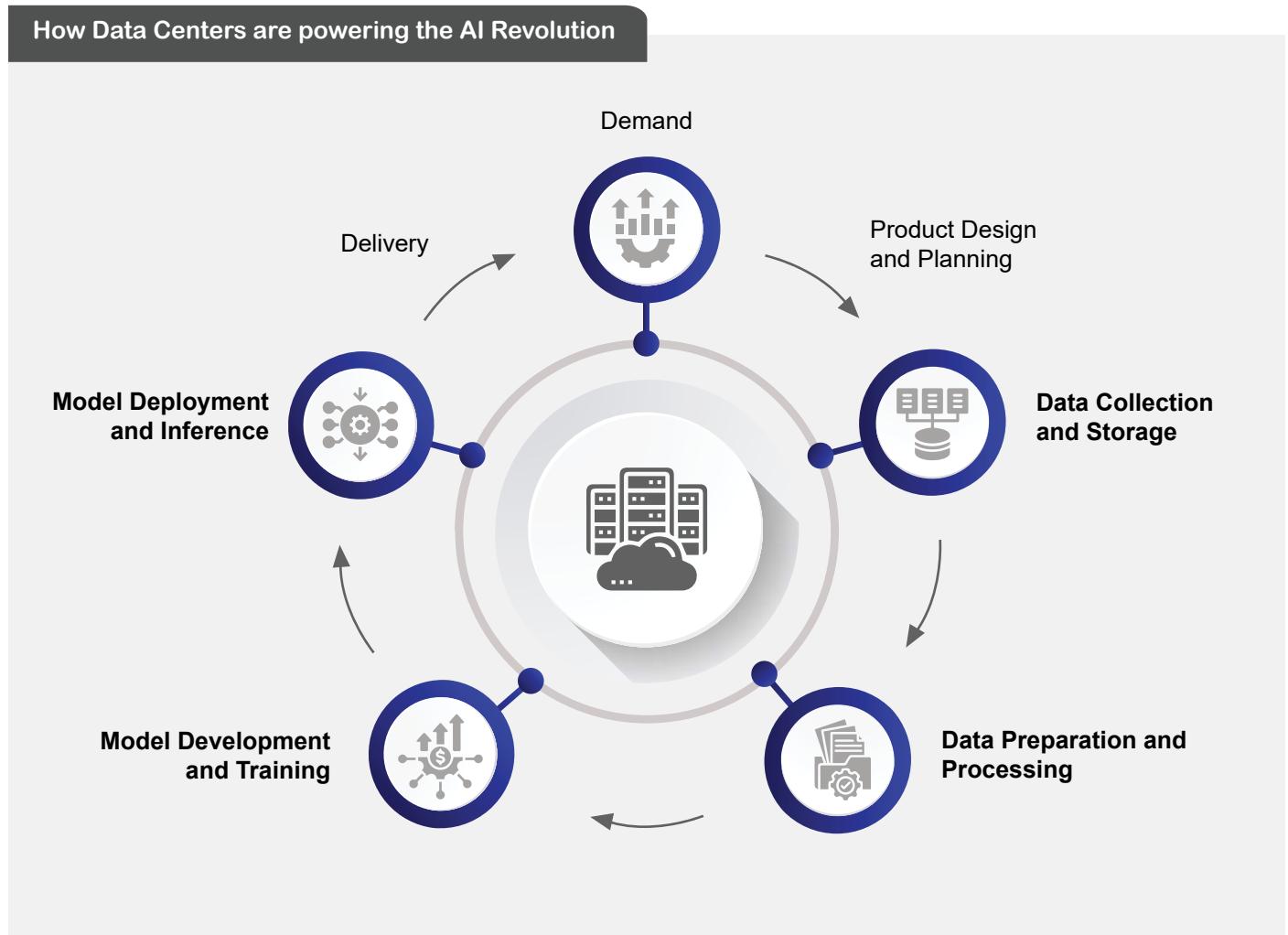
This section explores the critical forces shaping AI infrastructure—from the accelerating race to expand data centers to the looming energy crisis that could define the next phase of AI growth.



# Infrastructure

## Data Centers

### What You Need to Know: The Role of Data Centers in AI



AI development depends on vast computational infrastructure, and data centers are the backbone that powers every stage of the AI lifecycle—from data collection to real-world deployment.

- **Data Collection & Storage** – AI starts with data. Organisations gather vast amounts of structured and unstructured information from diverse sources, which must be securely stored and managed. Data centers provide the scalable storage infrastructure and high-speed access needed to house and retrieve this data efficiently.
- **Data Preparation & Processing** – Before AI can use data, it must be cleaned, labeled, and structured. This preprocessing happens in data centers, where distributed computing resources handle massive-scale data transformations, ensuring AI models are trained on high-quality, structured inputs.



- **Model Development & Training** – Training AI models requires immense computational power. Data centers house specialised AI accelerators (such as GPUs and TPUs) that enable deep learning frameworks to process trillions of operations per second, turning raw data into intelligent models.
- **Model Deployment & Inference** – Once trained, AI models must be deployed for real-world use, processing new inputs and generating responses. Inference happens in data centers, where optimised computing infrastructure ensures AI models operate with low latency and high throughput. Scaling AI in production is an essential challenge—as demand grows, data centers allow AI applications to serve millions of users simultaneously, maintaining reliability and efficiency across global markets.

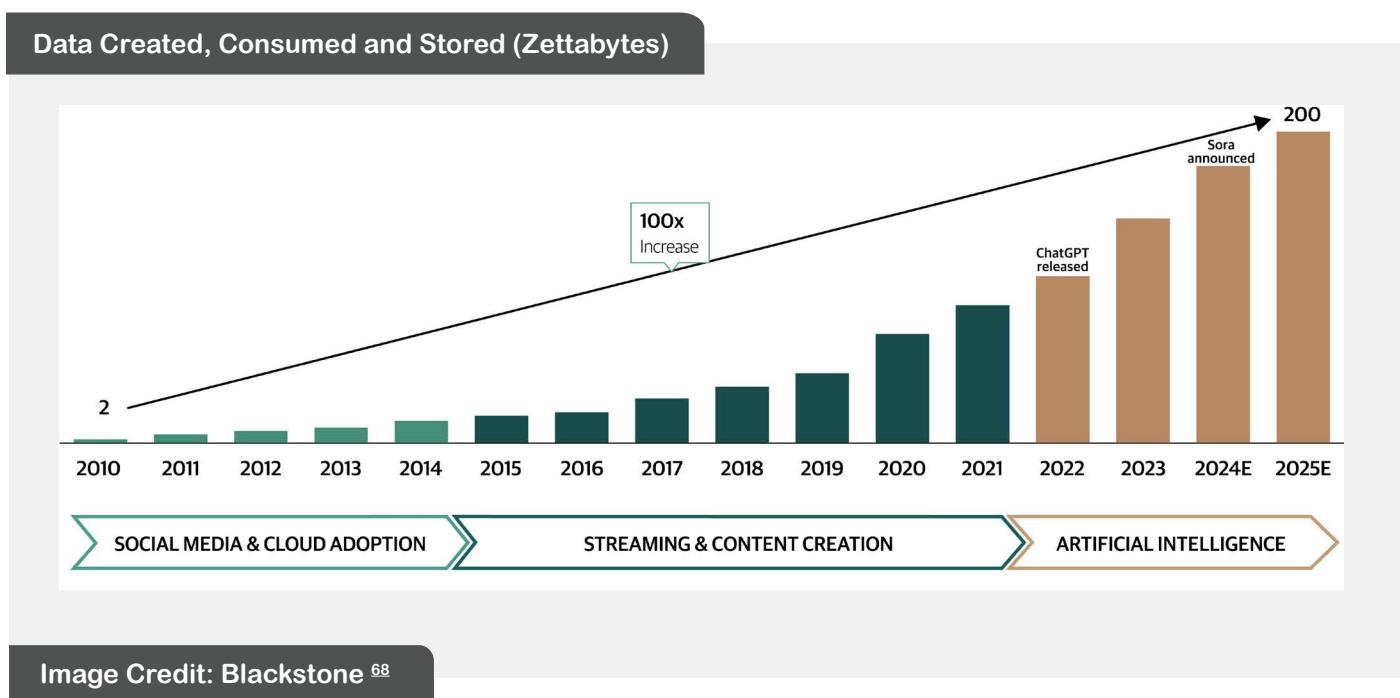
Data centers are **not just infrastructure**—they are the enabler of AI at every stage, determining how fast, scalable, and cost-effective AI solutions can be.

## What You Should Know: Demand and Supply for Data Centers

### Demand

The demand for AI-ready data centers is skyrocketing, driven by foundational model training, cloud-based inference, and enterprise AI adoption.

Data has always been the fuel of digital transformation, but AI is pushing data generation, consumption, and storage to unprecedented levels. Blackstone estimates that, **in 2025, AI will drive global data volumes to 100 times that of 2010**, matching or outpacing the internet, social media, and streaming in data intensity.



By 2028, Goldman Sachs expects **AI to represent about 19% of data center power demand.**<sup>69</sup> In response, an estimated **70% of total data center capacity will be optimised for AI workloads by 2030**, according to McKinsey.

Hyperscalers—Amazon, Microsoft, Google, Meta, and Baidu—are driving this demand. These companies are expanding data center capacity to support AI models like Gemini and ChatGPT. From 2023 to 2028, **hyperscalers will generate 60% of the industry's growth**, increasing their share of global data center power usage from 35% to 45%.

The current rate of **data center expansion is insufficient to meet AI demands**. Existing capacity is already exhausted: New facilities set to come online in the next 2–3 years are fully leased. In Northern Virginia, the world's largest data center hub, vacancy rates fell below 1% in 2024.<sup>70</sup>

## Supply

McKinsey estimates that to keep up with AI demand, the world must build twice the data center capacity developed since 2000—within just five years. **BCG expects \$1.8 trillion in new data center investments by 2030**. AI leaders, governments and investors are already committed to massive data center expansion:

### Hyperscalers

In many cases, hyperscalers are building the supply for their own demand:

- **Microsoft:** Announced an **\$80 billion** investment in AI-enabled data centers in 2025.<sup>71</sup>
- **Amazon:** Plans to invest “at least” **\$11 billion** in Georgia alone, to expand data center infrastructure.<sup>72</sup>
- **Meta:** Set to invest between **\$60 billion and \$65 billion** in 2025 to bolster its AI infrastructure, including constructing a data center housing **1.3 million Nvidia GPUs.**<sup>73</sup>
- **Alphabet (Google):** Has earmarked **\$75 billion** for capital expenditures in 2025, focusing on expanding its AI data centers to meet growing demand.<sup>74</sup>

## Governments

Recognising AI's strategic importance, governments are also making and facilitating significant investments:

- ▶ **United States:** The federal government launched the "Stargate Project," committing \$500 billion to AI infrastructure development.<sup>75</sup>
- ▶ **United Kingdom:** Announced £25 billion in new data center investments and plans to increase public sector computing capacity by 20-fold.<sup>76</sup>
- ▶ **France:** Unveiled a €109 billion investment plan to advance its AI sector, focusing on infrastructure development and computing clusters.<sup>77</sup>
- ▶ **United Arab Emirates:** Collaborated with France to invest between €30 billion and €50 billion in building a 1GW AI data center and other AI investments in Europe.<sup>78</sup>
- ▶ **Saudi Arabia:** Launched a \$100 billion AI initiative focused on building state-of-the-art data centers, supporting startups, and expanding AI infrastructure.<sup>79</sup>
- ▶ **China:** As of mid-2024, China invested approximately \$6.1 billion in a national project to construct computing data centers.<sup>80</sup>



## Investors

Investors are seizing opportunities in AI infrastructure:

- ▶ **Blackrock:** Formed the Global AI Infrastructure Investment Partnership (GAIIP) alongside Microsoft, Global Infrastructure Partners (GIP), and MGX, planning to raise \$80 billion to \$100 billion to construct data centers and the supporting grid energy infrastructure.<sup>81</sup>
- ▶ **Blackstone:** Has over \$70 billion in data center assets and an additional \$100 billion in their development pipeline.
- ▶ **Brookfield Asset Management:** Plans to invest billions in the AI sector and its necessary power infrastructure, including up to €20 billion already committed in France.<sup>82</sup>
- ▶ **Goodman Group:** Raised over \$4 billion to support a \$10 billion data center development program, which accounts for 46% of its projects.<sup>83</sup>
- ▶ **DigitalBridge:** Is one of the most active investors in AI data centers, managing over \$80 billion in infrastructure assets, including a leading role in Vantage Data Centers' \$9.2 billion equity investment round.<sup>84</sup>

## What Experts Are Talking About: The Barriers and Solutions to Data Center Growth

AI-driven data center expansion faces critical constraints that are shaping its future. While demand is skyrocketing, several bottlenecks—from energy costs to regulatory hurdles—threaten to slow growth. However, emerging solutions and strategic shifts are enabling expansion in new regions and accelerating deployment.

### Energy Costs & Grid Limitations

Power constraints are becoming a major bottleneck in data center hotspots like Northern Virginia and Santa Clara, CA. Utilities are struggling to build out transmission infrastructure quickly enough to meet demand, and concerns over electricity shortages are rising. Read more about AI's energy usage in the next section of this report.

In some cases, power is only allocated in small tranches (15–25MW per facility) while grid infrastructure is gradually expanded. In other cases, Government restrictions on grid access have halted new data centers altogether. In Ireland, for example, no new grid connections will be issued in Dublin until 2028 due to concerns that data centers will consume 28% of the country's electricity by 2031.<sup>85</sup>

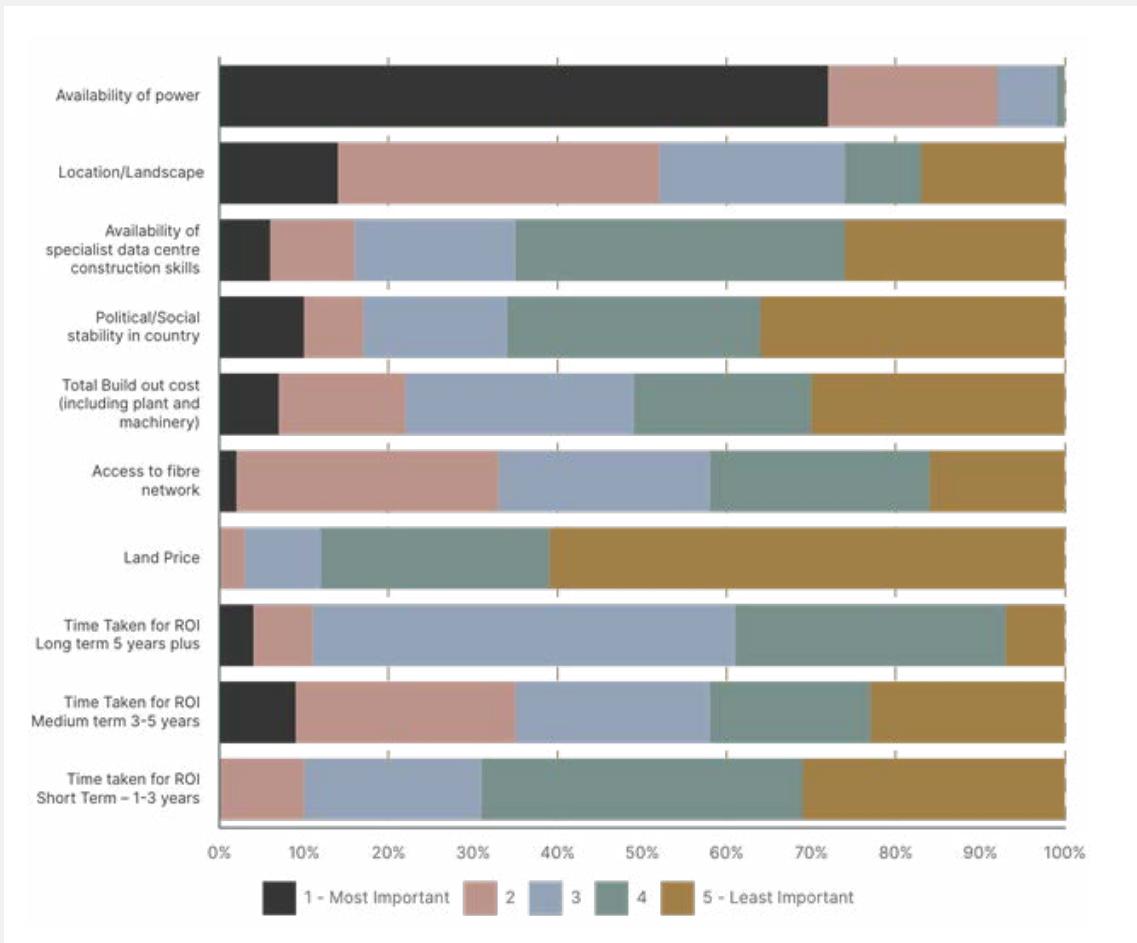
However, not all energy barriers are evenly distributed. While the U.S. still dominates the data center market—thanks to hyperscaler HQs, stable energy supply, and regulatory advantages—new regions are emerging as key growth areas:



- **Nordics:** A prime AI data center destination due to low-cost renewable energy and robust grid infrastructure.
- **Middle East (Saudi Arabia):** Attracting AI infrastructure investment due to abundant, low-cost electricity, as seen in Groq's \$1.5 billion data center expansion.
- **Remote U.S. states (Indiana, Iowa, Wyoming):** Chosen for AI model training due to lower energy costs, available land, and less grid congestion, making them ideal alternatives to overcrowded data center hubs.

Further, since the training phase of the AI lifecycle does not require close proximity to end users, training-focused data centers can be built where land and power are cheapest.

### Ranking Drivers of Data Center Choices



**Image Credit: BCS<sup>86</sup>**

(N=3000 senior data centre professionals across Europe, including owners, operators, developers, consultants and end users, surveyed by BCS Consultancy in 2024.)

While cost efficiency and market-driven resource allocation play a critical role in AI infrastructure expansion, **regulatory frameworks are also shaping where and how data centers are built**. **Data sovereignty laws**—such as the EU's GDPR and Indonesia's localisation mandates—are driving a regionalised approach to AI infrastructure, requiring companies to store and process data within specific jurisdictions.

Beyond sovereignty concerns, **permitting delays and environmental regulations** remain significant hurdles. Many sustainability policies now require data centers to be powered by renewable energy, adding complexity to site selection and construction timelines.

However, as AI capabilities become more strategic and geopolitical competition intensifies, many **governments are moving to ease regulatory and planning barriers** in order to accelerate data center expansion

## Regulatory Reforms in Europe

- ▶ Across Europe, governments are introducing planning reforms to cut through red tape and bureaucratic delays.<sup>87</sup> In the UK, for example, the government has pledged to "bulldoze through the ludicrous blockages" that have prevented AI-critical infrastructure from being built.
- ▶ A key shift has been the decision to classify data centers as Critical National Infrastructure (CNI)—a move that reflects their strategic importance to the economy and national security. This designation enhances government support for the sector, ensuring better preparedness against risks and threats.
- ▶ To streamline expansion, the UK government has proposed fast-tracking data center projects through the Nationally Significant Infrastructure Projects (NSIP) regime, which would allow planning approvals to go directly to the Secretary of State for decision rather than facing prolonged local processes.
- ▶ These reforms have been well received within the industry—**92%** of surveyed senior data center professionals in the 2024 BCS Consultancy study welcomed the decision to reform the planning process, while **93%** supported the move to classify data centers as CNI.
- ▶ As AI infrastructure becomes an economic and security priority, further regulatory shifts are expected worldwide, with **governments balancing sustainability goals with the need to remain competitive in AI development**.



## Construction Delays & Supply Chain Constraints

**Data center projects frequently exceed budgets and timelines,**<sup>88</sup> with supply chain disruptions affecting the availability of key electrical components and specialised labour. The rapid expansion of AI and cloud computing has intensified demand, creating bottlenecks for power, infrastructure materials, and skilled personnel, which in turn delays the completion of new facilities.<sup>89</sup>

**A shortage of qualified data center professionals** is emerging as a critical challenge. In Europe, industry leaders report growing concerns over the availability of skilled design, construction, and operational staff. BCS' survey found that nearly all respondents **expect a decline in available qualified personnel by 2025**, even as demand rises. Recruiting and retaining specialists is particularly difficult in remote regions, where major data center hubs are expanding.

### Survey Responses: 'It is increasingly difficult to source sufficiently skilled professionals to design/build/operate data centers in Europe'

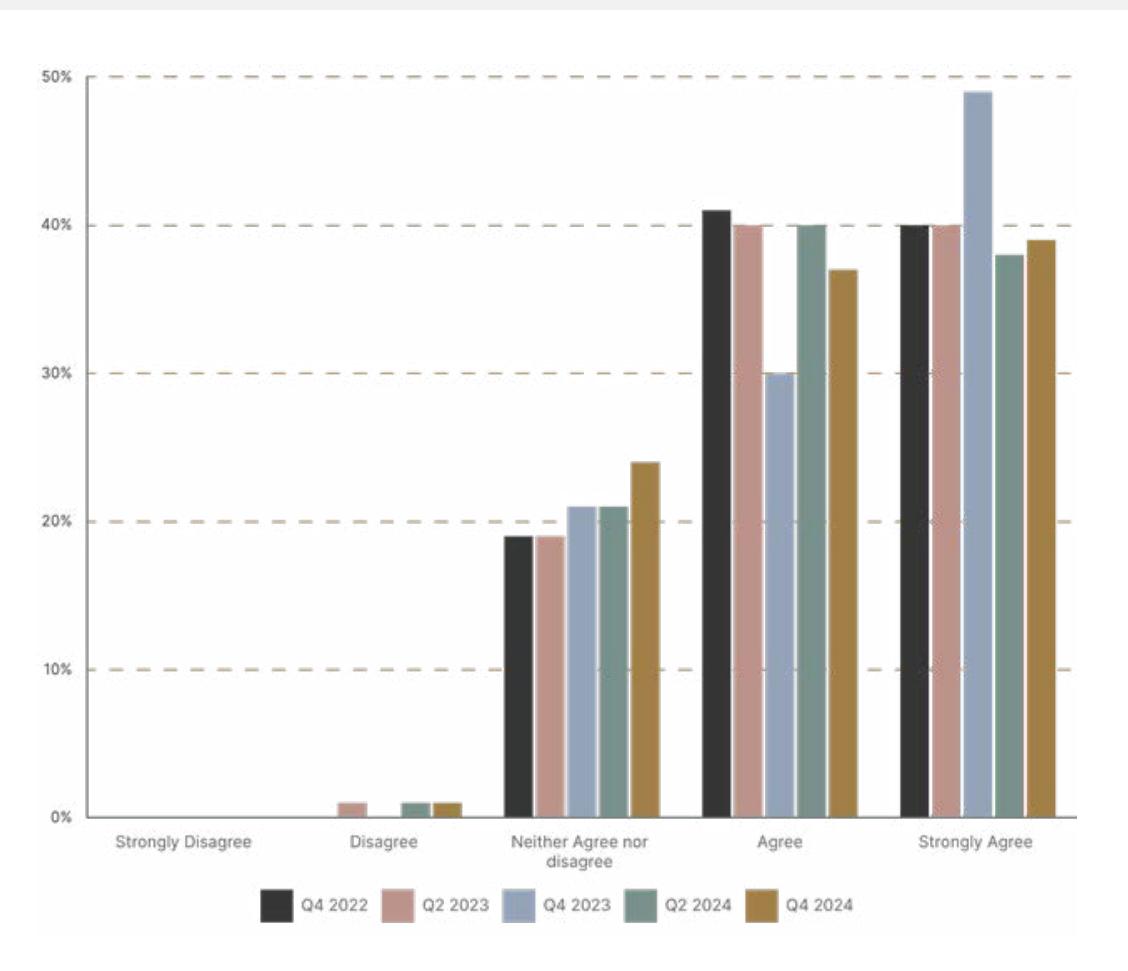


Image Credit: BCS

Some companies are addressing this gap through workforce development programs. For example, Microsoft has partnered with local community colleges to launch 20+ global data center academies, training the next generation of AI infrastructure specialists.

In other cases, **companies are making data center construction more efficient with technology**. In particular, inefficiencies in project management lead to significant, and frequent, cost overruns. Inconsistent reporting, siloed systems, and slow decision-making cycles make it difficult to identify risks early and keep projects on track.

Foresight, for example, has developed an AI-powered project management platform designed to streamline workflows, mitigate risks, and enhance efficiency in large-scale data center construction.<sup>90</sup> Their platform automates, facilitates and centralises the project management workflow:

- **Task Prioritisation:** Uses AI to identify critical tasks, flag milestone blockers, and create structured completion plans.

---

- **Scheduling & Reporting:** Enhances schedule transparency and forecasting with automated reports, custom dashboards, and trend analysis, improving visibility and decision-making.

---

- **Risk Identification:** Provides real-time risk analysis, logic quality checks, and DCMA assessments to detect bottlenecks early.

---

- **Collaboration & Stakeholder Integration:** Centralises project workflows by assigning tasks, automating notifications, and integrating external schedules, ensuring all stakeholders stay aligned.

---

- **Continuous Learning & optimisation:** Leverages historical project data to forecast activity durations, assess network impacts of overruns, and track phase performance trends.

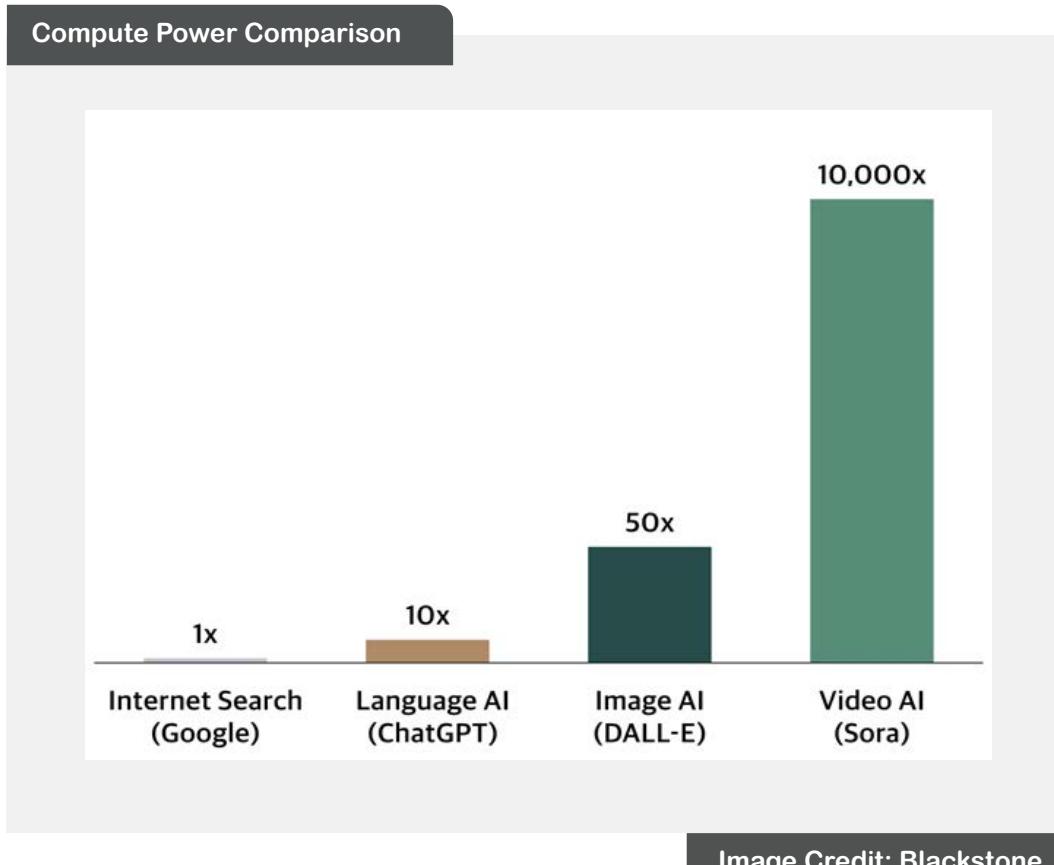
Industry estimates suggest that, even if all current construction plans are completed on time, the U.S. will still face a **data center capacity deficit of over 15 GW by 2030**.<sup>91</sup> In an industry where delays are costly and demand is relentless, advanced planning, automation, and risk mitigation strategies will be critical for meeting the next wave of AI-driven infrastructure needs.



# Energy

## Powering AI

### What You Need to Know: Why Does AI Require So Much Energy?



AI is fundamentally a computation-heavy technology. Computation itself is simply the act of processing data—turning inputs into meaningful outputs using mathematical operations. The more complex the task, the greater the computational power required and the higher the energy demand.

Unlike conventional applications, AI workloads—especially training large models—demand an extraordinary level of computational intensity. This is due to both the sheer volume of data AI systems must process and the iterative nature of training, where models continuously refine their predictions by running billions of calculations across specialised chips.

In 2023, training GPT-4 was estimated to use computation comparable to the combined power of running millions of Frontier supercomputers<sup>92</sup> or performing more than 60 times the computation required to train GPT-3.<sup>93</sup> The energy needed to sustain this process is significant, and as AI models grow more sophisticated, the demands on infrastructure and power sources are only increasing.

Each AI model runs on high-performance computing hardware—most commonly GPUs (graphics processing units) and TPUs (tensor processing units)—which require immense amounts of electricity to operate. These chips perform parallel computations at high speeds but generate substantial heat as a byproduct. As workloads scale, so do the energy needs—not just to power computation, but also to manage the resulting heat and ensure system stability.

## Hence, AI demands energy in multiple forms:

### ➤ Electricity

The foundation of AI infrastructure. Every stage of the AI lifecycle—data storage, processing, model training, and inference—depends on high-power compute clusters. The most energy-intensive phase is training, where GPUs and TPUs consume tens of megawatts per deployment. Once models are trained, they enter the inference phase, where they continuously process new data in real time, drawing lower but sustained power.

### ➤ Water & Air

The extreme heat generated by AI chips must be managed to avoid hardware failure. Traditional air cooling uses fans and air conditioning, but as power densities rise, water-cooled systems are becoming more common due to their greater efficiency in high-density environments. Some facilities rely on evaporative cooling, which consumes large volumes of water, while others use liquid cooling loops that cycle water through specialised cooling units.

### ➤ Diesel & Batteries

To prevent service disruptions, AI data centers have extensive backup systems, including industrial-scale batteries and diesel generators that can sustain operations during grid failures.

### ➤ Land & Physical Infrastructure

AI-ready data centers require more than just power—they need vast amounts of space to house racks of servers, cooling equipment, and power distribution systems. As power densities increase, facilities are shifting toward vertical expansion (multi-story data centers) and modular architectures that can scale more flexibly.

Infrastructure providers and technology firms are racing to balance AI performance with efficiency. As adoption accelerates, power constraints will become a defining challenge, forcing innovation in both hardware design and energy sourcing. The next subsection will explore exactly how much energy AI is consuming today and where the biggest pressures lie.

## What You Need to Know: How Much Energy Does AI Actually Use?

While data centers have long been a backbone of the digital economy, the rapid expansion of AI workloads has dramatically increased their power requirements. Our datasheet provides a breakdown: <sup>94</sup> <sup>95</sup>



## How Much Power Does AI Actually Use?

A single ChatGPT query requires 2.9 watt-hours of electricity



10x more than a Google Search



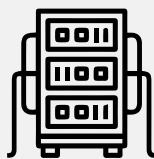
AI-ready data centers require significantly more power than their predecessors:

### A Single Power Rack



On Average

Equivalent to **four U.S. households.**  
(17 kW)



For Advanced Training

Equivalent to an **entire data center** from the 2010s  
(120 kW)



Hyperscalers operate data centers with tens of thousands of racks



Equivalent to a **small**

300MW



In 2023 alone, Google's data centers consumed 15.5 TWh of electricity-equivalent to the annual energy consumption of ~1.45 million U.S. households.

AI's energy consumption is already significantly impacting nations

In 2022, data centers accounted for nearly  
**20% of Ireland's total electricity use**



In 2024 alone, 5,000 megawatts (MW) of data center capacity was added in the U.S.- **roughly 1% of the nation's total power consumption.**

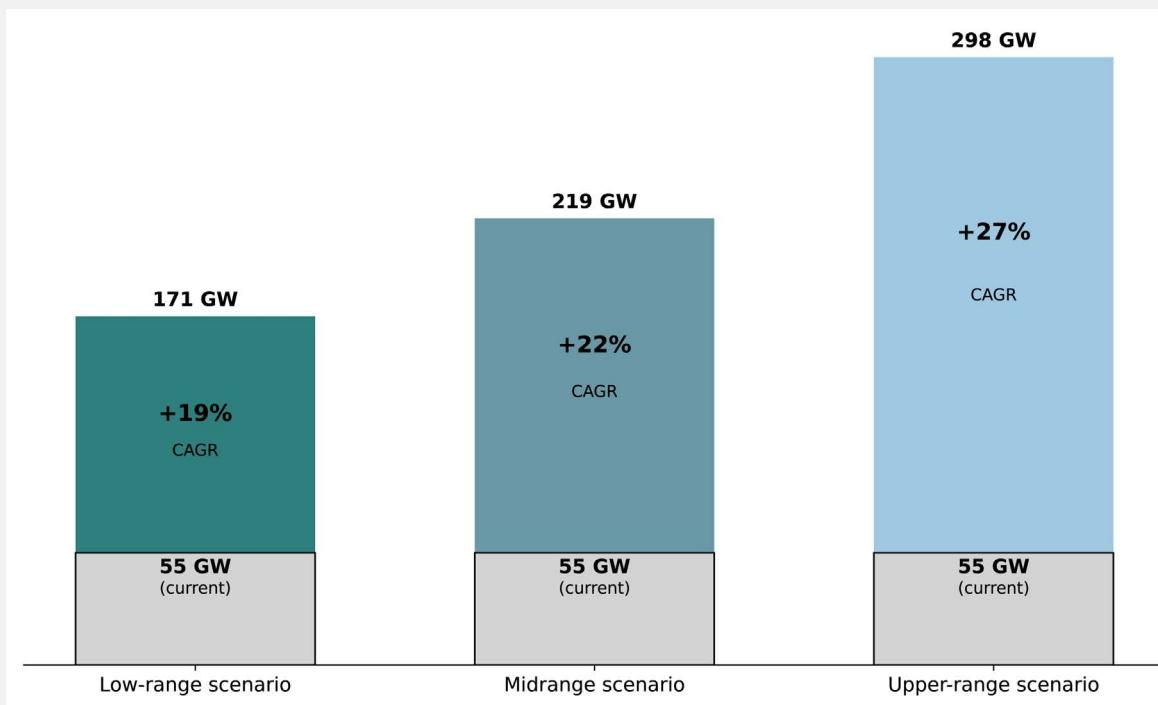
## The Growth of AI Power Usage

A decade ago, a 30 MW data center was considered large—today, 200 MW is common, and 500 MW facilities are under construction.<sup>96</sup> This growth is only set to continue:

- ▶ **Data Center Clusters:** OpenAI CEO Sam Altman has proposed building 5,000 MW AI data center clusters across the U.S.—each equivalent to all U.S. data center capacity added in 2024.<sup>97</sup>
- ▶ **US:** By 2030, U.S. data center electricity demand is expected to rise from 3% of total power usage (2022) to 8%, making data centers the largest single driver of U.S. electricity demand growth.<sup>98</sup>
- ▶ **Europe:** Europe's power demand could grow by 40–50% between 2023 and 2033 due to AI and broader electrification efforts. By 2030, the energy consumption of European data centers will match the current total power use of Portugal, Greece, and the Netherlands combined.<sup>99</sup>
- ▶ **Total Data Center Capacity:** Global data center capacity demand will rise from 60 GW today to between 171 and 219 GW by 2030—with a high-end estimate of 298 GW.<sup>100</sup>

GenAI's iterative nature, contextual awareness, and non-deterministic outputs introduce new risks that traditional validation frameworks cannot fully address, requiring expanded, automation-driven testing.

### Three Scenarios: McKinsey's demand for data centre capacity forecasts, (gigawatts)



The question is no longer whether AI will reshape energy consumption—but whether the energy supply can keep up. We explore this question in the [next section](#).



## What Experts Are Talking About: How Are We Going to Fuel the AI Revolution?



Rising energy demand over the next five years, fueled in part by the data center boom, will drive the largest five-year expansion of energy capacity in history

— BCG, 2025



According to John Pettigrew, Chief Executive of the UK's National Grid, data center power usage in the UK alone is on track to increase six-fold in the next decade. "The grid is becoming constrained," he warned, noting that "bold action" will be needed to create a network able to cope with "dramatically" growing demand.<sup>101</sup>

Europe's challenges are especially pronounced. Possessing some of the world's oldest power grids, the continent faces urgent modernisation needs. Between 2023 and 2033, it is projected that Europe will need **nearly €1.6 trillion** in combined grid and renewable energy investment to keep pace with AI-driven demand.<sup>102</sup> Approximately **€800 billion** (about \$861 billion) is earmarked for transmission and distribution upgrades, with another **€850 billion** for new wind (both onshore and offshore) and solar capacity. Such expenditures aim not only to power AI data centers but also to meet broader electrification goals as countries strive for net-zero emissions by 2050.

### Hopes—and Pressures—for Renewables

Grid operators face mounting pressure to upgrade infrastructure in a way that not only meets rising electricity demand but also integrates sustainable energy sources.

However, the urgency to power AI sustainably has placed immense strain on renewable energy solutions. Data center operators are taking foundational steps to reduce their carbon footprints, including purchasing renewable energy credits and signing Power Purchase Agreements (PPAs)—long-term contracts between energy buyers and renewable power producers that secure stable electricity pricing and incentivise clean energy projects.

But the current limitations surrounding renewable energy leave key questions unanswered:

- **Intermittency:** Wind and solar power depend on weather conditions, meaning their output fluctuates and cannot always meet continuous power demands.
- **Capacity:** The capacity factor—the percentage of actual output compared to maximum potential output—is around 30% for wind and solar,<sup>103</sup> significantly lower than fossil fuels or nuclear, which operate closer to 90%.<sup>104</sup>
- **Storage:** Battery technology offers a potential solution for balancing renewable supply and demand, but current systems remain expensive and limited in scale, requiring further advancements in efficiency and grid integration.



## The Push Toward Nuclear

Nuclear energy is emerging as an attractive, carbon-free alternative for AI data center operators. Amazon, for example, has an agreement with Talen Energy to power one of its data centers using nuclear energy—part of the tech giant's effort to reach net-zero carbon by 2040.<sup>105</sup> In the UK, the government is exploring partnerships with nuclear providers and the development of Small Modular Reactors (SMRs) to power AI "Growth Zones."<sup>106</sup> Under these plans, a new data center pilot would begin at **100 MW** of capacity—with plans to scale to **500 MW**—drawing on dedicated, clean energy sources. An "AI Energy Council," co-chaired by the UK's Science and Technology Secretary and Energy Secretary, aims to accelerate investment in renewables, nuclear, and other innovative energy solutions, while also looking at how AI itself can optimise energy systems.

Nuclear's appeal in this context lies in its reliability and minimal carbon footprint. For AI workloads that require consistent, high-density power, nuclear can complement wind and solar while sidestepping the carbon emissions of fossil fuels. Still, cost overruns, lengthy construction times, and public concerns about nuclear safety remain obstacles that policymakers and industry leaders must navigate.

## The (Continuing) Role of Fossil Fuels

Despite a global push for clean energy, fossil fuels remain a significant factor in powering AI. In the United States, gas-fired power plants are proliferating—up to **80 new facilities by 2030**, adding 46 GW of capacity (equivalent to the entire electricity system of Norway).<sup>107</sup> In December 2022, Entergy announced a **\$3.2 billion** plan to build three gas plants totaling 2.3 GW, largely to serve a **\$10 billion** AI data center owned by Meta. Meanwhile, major oil producers like ExxonMobil and Chevron are exploring ways to supply AI data centers directly, bypassing the grid altogether.

The use of fossil fuels also carries political dimensions. In the U.S., President Donald Trump has advocated for increased domestic energy production to reduce reliance on China—a contention running perfectly parallel to the nations' ongoing AI arms race, brewing what could be the perfect storm.

## Finding Efficiencies

Energy is going to remain an important part of the AI conversation. However the extent to which it is critical, or even limiting to progress, also depends on our progress towards greater efficiency in running AI systems. A few key avenues stand out:

- **Optimised Cooling Systems** – Transitioning from traditional air cooling to liquid cooling significantly reduces energy waste and improves power usage effectiveness (PUE), cutting operational costs. Some data centers have reported up to a **10% improvement in PUE** with liquid cooling compared to air-based systems.
- **Economies of Scale in Power Distribution** – Deploying larger switchgear, high-capacity power distribution units, and upgrading to 48-volt server power supplies reduces energy loss, infrastructure complexity, and maintenance costs.
- **Smarter Workload Management** – Balancing AI computation between cloud and edge computing reduces data center load, optimising power consumption at scale.
- **Algorithmic Efficiency** – Advances like DeepSeek's breakthrough demonstrate how optimised inference and training techniques can dramatically cut energy use while maintaining cutting-edge AI performance.



- **AI-Optimised Semiconductors –** Companies like NVIDIA, AMD,<sup>108</sup> and Intel<sup>109</sup> are investing heavily in developing specialised AI chips to enhance performance and energy efficiency. Additionally, technologies such as quantum are converging on AI use cases. For instance, Google's "Willow" quantum chip, introduced in December 2024, features 105 qubits and has achieved substantial error reduction, marking a pivotal step toward practical quantum computing applications.

NVIDIA is adopting a holistic approach to creating efficiencies in AI by offering **integrated packages that encompass hardware, software, and support services**. A notable example is the NVIDIA DGX B200 system, a unified platform designed to streamline AI workflows from development to deployment, which combines eight NVIDIA Blackwell GPUs with comprehensive software stacks such as NVIDIA AI Enterprise and NVIDIA Base Command. Their specialised AI server runs at **15x the speed of today's server for only 2x the power**.

Since its release in 2024, the package has already gained significant traction. The University of Florida investing \$24 million in a supercomputer comprising 63 DGX B200 systems,<sup>110</sup> and the Internal Revenue Service (IRS) acquiring a DGX SuperPOD with 31 DGX B200 units to enhance its data analytics capabilities.<sup>111</sup> Additionally, Microsoft Azure became the first cloud service to deploy NVIDIA's Blackwell system, integrating GB200-powered AI servers into its infrastructure.<sup>112</sup> SoftBank is also partnering with NVIDIA to deploy a GB200-based AI supercomputer in Japan, aiming to support AI applications and accelerate the development of AI-powered 5G networks.<sup>113</sup>

## Conclusion

The AI revolution is still in its early phases—defined by excitement, rapid progress, and inevitable imperfections. Successful leaders are not chasing AI's limitless potential but instead focusing on a small number of high-impact areas, optimising heavily, and applying their learnings to future initiatives.

Through our analysis, we have established that:

➤ **Technology is ready to deliver value:**

While challenges remain, AI capabilities are advancing rapidly. Organisations are moving beyond experimental adoption, deploying AI in ways that improve automation, efficacy, and safety. Custom models, RAG, and agent-led workflows are reshaping enterprise AI, enabling businesses to tailor solutions to their specific needs.

➤ **People are ready to embrace change:**

Contrary to cautious executive assumptions, employees are more prepared for AI-driven change than expected.<sup>114</sup> A growing body of research underscores that workforce empowerment—rather than displacement—is the key to unlocking AI's full potential. However, more needs to be done to upskill the workforce on emerging AI capabilities, especially in risk management.

➤ **Businesses are in a race:**

The scale and complexity of modern enterprises mean that AI adoption is progressing at different speeds. However, speed alone is not enough—the winners will be those that adopt AI intelligently, balancing technical advancements with regulatory compliance and strategic implementation. Given the depth of expertise required across technology, compliance, and security, **many firms are turning to AI services as a bridge between legacy models and AI-driven operations**.



# Services as a Bridge

The demand for AI consulting and implementation services is surging, with IBM WatsonX generating a \$1 billion book of business and AI-focused consulting revenue doubling year-over-year. SaaS providers like LucidWorks and Invoke.ai are integrating professional services to guide AI rollouts,<sup>115</sup> while firms like Accenture saw record bookings of \$3 billion for GenAI consulting in 2024.<sup>116</sup>

However, it's not only the large incumbents who stand to gain. A new breed of AI-focused service providers is emerging—lean, highly specialised, and built to maximise AI's value while avoiding the pitfalls of traditional consulting models. These firms are capitalising on the shake-up within legacy consulting, where outdated business models, rigid hierarchies, and misaligned incentives have left gaps in the market.

**Three growing companies have caught our eye:**

**ValidMind is performing Model Risk Management-as-a-Service (MRMaaS).**<sup>117</sup> MRMaaS is an outcome-based approach that combines its specialised SaaS platform with expert-driven validation services. This integrated model streamlines AI validation by leveraging ValidMind's technology alongside a strategic service component, where its experts design and execute tailored validation strategies. Early pilots have delivered significant cost savings and faster validation cycles, eliminating the need for banks to hire specialised staff or engage large consultancies.



ValidMind's workflow and communication tools gave us a worry-free experience, allowing us to focus on value-driven aspects of model validation.

— GBC Chief Risk Officer, Adam Ennami



**Quantum Rise is rethinking AI consulting from the ground up.** Rather than charging by the hour, they focus on outcome-based pricing, ensuring incentives are aligned with their clients' success. Their approach blends AI due diligence, integration, and transformation strategy, helping businesses adopt AI efficiently and iteratively. Importantly, they position AI as a long-term asset rather than a one-off implementation, offering reusable solutions that scale across an enterprise. Quantum Rise also provides AI training, helping organisations build internal capabilities instead of creating long-term reliance on external consultants.



The pace of change of AI can be overwhelming. We see our role as being a filter for some of that, and the noise around AI and automation. We focus on where the money is, either in savings or in net new revenue—and work back from there to fit the right technology to make it work.

— Alex Kelleher, CEO & Founder, Quantum Rise



**AI & Partners specialises in compliance and governance, particularly around the EU AI Act**—widely regarded as the global benchmark for AI regulation. As regulation becomes a key determinant of AI adoption, AI & Partners helps organisations navigate complex compliance requirements through AI impact assessments, ISO 42001 pre-audits, and legal frameworks. Their structured, step-by-step approach accelerates time-to-compliance while reducing legal and operational risks. With governments and enterprises increasingly requiring AI governance transparency, AI & Partners is positioned as a critical ally in building responsible, trustworthy AI systems.



The EU AI Act signals a new era of accountability in AI-driven compliance. As generative AI reshapes risk management, firms must align innovation with regulation—ensuring transparency, fairness, and control. This act isn't just a restriction; it's a blueprint for trust in AI-powered decision-making.

— Sean Musch, CEO & Founder, AI & Partners



## Stay Ahead in The AI Revolution

For enterprises, startups, and policymakers seeking clarity in this fast-moving space, Parker & Lawrence's research provides deep insight into AI, risk and compliance. Whether helping vendors position their technology, enterprises identify opportunities, or governments refine AI strategy, our mission remains the same: enabling the responsible adoption of AI at scale.

**The AI revolution is here. The question is no longer whether to adapt, but how to lead.**

# Where are we in the AI Revolution?



## Adoption

Generative AI's early adoption has been greater than computers or the internet.

**72%** of organisations have adopted AI, with **65%** reporting regular use of generative AI.

**4%** of organisations report zero use of generative AI.

**32.5%** of businesses are already using fine-tuned models and **64%** are developing custom applications.

**51%** of organisations have implemented AI agents in production environments.

**32%** of the US population aged 18-64 say they have used generative AI at least once in the week prior.

**37%** of consumers are already comfortable with AI agents creating more personalised content for them.



## Impact

These early stages are defined as much by struggle as they are by excitement. Though leaders are pulling away.

**74%** of companies are not yet achieving value at scale with AI.

**49%** of respondents claim generative AI use cases are not delivering the value anticipated.

Organisations leading in AI adoption are realising:

- **50%** higher revenue growth
- **40%** higher return on invested capital
- **40%** higher customer satisfaction

**92%** of service teams with AI say it reduces their costs.

**\$45 billion** revenue generated by GenAI in 2024.

**75%** of generative AI's current economic impact comes from customer operations, marketing & sales, software engineering, and R&D.



## Plans

The scale of AI investment plans are so large that today's achievements will soon be dwarfed.

**94%** of corporate leaders see AI significantly influencing operational and product strategies.

**92%** of organisations plan to increase AI investments over the next three years.

**33%** of Companies Plan to Spend More than **\$25 Million** On AI in 2025.

**\$500 billion** committed to AI infrastructure development by the US federal government.

**\$100 billion** Saudi government initiative focused on supporting AI startups and expanding AI infrastructure.

**20x** increase in the UK's Artificial Intelligence Research Resource by 2030, according to government plans.



## Projections

AI innovation is widely believed to fundamentally change the way we operate as individuals, institutions and economies.

**\$17.1 trillion-\$25.6 trillion** added to the global economy annually by 2040 by AI.

**60-70%** of work *could* be automated with today's models, **50%** *will* be automated between 2030-2060.

**75%** of professionals predict significant or disruptive change in their industries due to generative AI.

**25-40%** of enterprises will adopt RAG by 2030.

**25%** of enterprises that use generative AI will deploy agents in 2025, increasing to **50%** by 2027.

**19%** of data center power will be consumed by AI in 2028, and **70%** of data center capacity will be optimised for AI by 2030.

